Propriétés théoriques et applications en statistique et en simulation de processus et de champs aléatoires stationnaires

Lionel Truquet

UNIVERSITÉ PARIS I PANTHÉON-SORBONNE
U.F.R. MATHÉMATIQUES

<u>THÈSE</u>

pour obtenir le grade de
DOCTEUR ÈS-SCIENCES
SPÉCIALITÉ MATHÉMATIQUES APPLIQUÉES

présentée et soutenue publiquement par

Lionel TRUQUET

le 10 décembre 2008

sous le titre

PROPRIETES THEORIQUES ET APPLICATIONS EN STATISTIQUE ET EN SIMULATION DE PROCESSUS ET DE CHAMPS ALEATOIRES STATIONNAIRES

Directeurs de Thèse

M. Paul DOUKHAN, Professeur à l'Université de Cergy Pontoise,
M. Jean-Marc BARDET, Professeur à l'Université Paris 1 Panthéon-Sorbonne.

Jury

Rapporteurs :	M. Bernard DELYON, Professeur à l'Université de Rennes 1.
	M. Piotr KOKOSZKA, Professeur à l'Université de Utah.
	M. Qi-Man SHAO, Professeur à l'Université de Hong Kong.
Examinateurs :	Youri DAVYDOV, Professeur à l'Université Lille 1.
	Xavier GUYON, Professeur émérite à l'Université Paris 1.
	Alain LATOUR, Maître de Conférences à l'Université de Grenoble.
	Florence MERLEVEDE, Professeur à l'Université de Marne la Vallée.
	Jean-Michel ZAKOIAN, Professeur à l'Université Lille 3.

Remerciements

Voici enfin une page sans mathématiques. L'occasion de remercier et saluer ceux qui ont aidé au bon déroulement de ce travail, par leur réflexion, leurs remarques, ou juste leur présence.

Je tiens à remercier Paul Doukhan, qui par sa disponibilité tout au long de cette thèse, aura permis de nombreuses discussions sur des thèmes mathématiques variés. Je remercie également Jean-Marc Bardet qui m'a aidé pour la mise en forme des derniers sujets.

L'équipe du SAMOS (Omar Aboura, Mouhamad Allaya, Charles Bouveyron, Camille Brunet, Marie Cottrell, Faniaha Dimby, Khalifa Es-Sebaiy, Patrice Gaubert, Cécile Hardouin, Aurélien Hazan, Imen Kammoun, Annie Millet, Madalina Olteanu, Joseph Rynkiewicz, Ciprian Tudor) a toujours su se montrer accueillante durant cette période, quelles que soient les circonstances.

J'ai aussi partagé le quotidien de nombreux doctorants, chargés de cours et professeurs de l'Ensae (Pierre Alquier, Patrice Bertail, Jean-François Chassagneux, Mathieu Cornec, Xavier d'Haultfoeuille, Romuald Elie, Emmanuelle Gautherat, Eric Gautier, Ghislaine Gayraud, Aude Grelaud, Hugo Harari, Olivier Lopez, Karim Lounici, Xavier Mary, Christian-Yann Robert, Christian Robert, Judith Rousseau). Nous avons échangé des travaux dirigés, parfois des inquiétudes, la bonne humeur de chacun restera un bon souvenir même camouflée dans cet étrange bureau F14 au fond de l'Insee.

Mes collaborateurs ont été là aussi pour faire évoluer la mise en forme des idées, Alain Latour, Nathanaël Mayo et Olivier Wintenberger.

Je remercie les rapporteurs Bernard Delyon, Piotr Kokoszka et Qi-Man Shao pour leur analyse rigoureuse du contenu de ce manuscrit, ainsi que Youri Davydov, Xavier Guyon, Alain Latour, Florence Merlevède et Jean-Michel Zakoïan, pour leur participation en tant que membres du jury.

Enfin, je ne peux finir sans un mot et une pensée pour l'attention et la patience que m'a accordées Anaïs, quelles que soient les difficultés.

Résumé

Ce travail doctoral étudie les propriétés théoriques et asymptotiques des processus et des champs aléatoires stationnaires dont se déduisent des applications en statistique et en simulation. Une première partie (Chapitres 2, 3 et 4) a pour objectif de construire des nouveaux modèles de champs aléatoires de type autorégressifs, sous forme de schémas de Bernoulli, et de donner des résultats au sujet de leur théorie limite. Des notions de dépendance faible sont utilisées, plus générale que les notions bien connues de mélange fort ou d'association. Nous envisagerons un principe d'invariance, faible et fort, pour les champs aléatoires considérés.
Dans un deuxième temps, nous nous intéressons à quelques problèmes d'estimation dans deux contextes de dépendance bien précis. Nous étudions au Chapitre 5 un problème de simulation de textures dans un contexte de rééchantillonnage pour des champs de Markov fortement mélangeants dans un cadre non paramétrique. Le Chapitre 6 est consacré à la construction et à l'estimation des paramètres d'une nouvelle série chronologique à valeurs entières de type ARCH. La construction est établie en utilisant des arguments de contraction établis dans le cadre des champs aléatoires et le comportement asymptotique des estimateurs des paramètres, obtenus par quasi-maximum de vraisemblance gaussien est fondée sur des arguments de type différence de martingales. Enfin nous présentons au Chapitre 7 une nouvelle méthode d'estimation des paramètres pour des modèles ARCH de type markoviens, méthode obtenue en lissant la quasi vraisemblance gaussienne et nous appliquons cette méthode à une série hétéroscedastique de type LARCH pour laquelle les faibles valeurs de la variance conditionnelle rendent difficile l'utilisation de la méthode classique du quasi maximum de vraisemblance.

Abstract

This PhD thesis studies theorical and asymptotic properties of processes and random fields with some applications in statistics and simulation. A first part (Chapter 2, 3 and 4) is devoted to the construction of new models of random fields with a random error, expressed in term of Bernoulli shifts and to give some results about their limit theory. Weak dependence conditions used are proved to be more general than the well known notions such as strong mixing or association. We will study in this part the weak and strong invariance principle, for the random fields of interest.
The second part of this thesis will be devoted to study estimation and simulation's problems with two kinds of dependence contexts. In Chapter 5, we first consider the question of texture simulations, with a non parametric resampling scheme for strong mixing random fields. The Chapter 6 is devoted to the construction and the parametric estimation of a new integer valued ARCH time serie. The existence result uses contraction arguments etablished in the first part for random fields and the asymptotic behaviour of parameters estimators, obtained using the (Gaussian) Quasi Maximum Likelihood Estimator (QMLE), is etablished with martingal differences type arguments. Finally, in Chapter 7, we introduce a new estimation procedure for Markovian ARCH models. The principle of this method is to smooth the Gaussian QML. We apply this method to the parametric estimation of LARCH type processes, for which the small values of the conditional variance make difficult to apply the usual QMLE technique.

Table des matières

Chapitre 1

Synthèse des travaux

1.1 Dépendance faible des champs aléatoires

Plusieurs mesures de dépendance ont été introduites pour modéliser les champs aléatoires. Il s'agit bien souvent de méthodes d'abord introduites dans le cadre des séries temporelles puis adaptées au cas des champs aléatoires. La mesure de dépendance la plus connue est le mélange fort ($\alpha-$mélangeant) et a été introduite à l'origine par Rosenblatt (1956) [30]. Nous référons au livre de Doukhan (1994) [10] pour la définition et les exemples d'autres types de mélange ainsi que de nombreux théorèmes limites très utiles. Nous rappelons la définition des coefficients de $\alpha-$mélange. Soit r une distance sur $\mathbb{R}^d$, $d \in \mathbb{N}^*$.

Definition 1.1 *Si X est un champ aléatoire indexé par $\mathbb{Z}^d$, on définit pour $a, b \in \mathbb{N}^* \cup \{\infty\}$ et $k \in \mathbb{R}_+$:*

$$\alpha_X(k, a, b) = \sup_{E,F \in \mathbb{Z}^d, r(E,F)=k, |E|=a, |F|=b} \big\{|P(A \cap B) - P(A)P(B)|,\ A \in \sigma(X_E), B \in \sigma(X_F)\big\},$$

où pour $A \subset \mathbb{Z}^d$, $X_A = (X_j)_{j \in A}$. Le champ aléatoire X est dit $\alpha-$mélangeant si :

$$\lim_{k \to \infty} \alpha_X(k, a, b) = 0, \quad a, b \in \mathbb{N}^*.$$

Nous rappelons aussi la définition de l'association pour les champs aléatoires. Cette notion est introduite par Esary, Proschan, et Walkup [17]. Une référence récente pour les théorèmes limites de ces champs ainsi que de certaines de leurs fonctionnelles est le livre de Bulinski et Shaskin [7].

Definition 1.2 *Un champ aléatoire X indexé $\mathbb{Z}^d$ à valeurs réelles est dit associé si pour toute partie I finie de $\mathbb{Z}^d$ et pour toutes fonctions f, g définies sur $\mathbb{R}^I$, bornées et croissantes coordonnée par coordonnée :*

$$Cov(f(X_I), g(X_I)) \geq 0,$$

où $X_I = (X_i)_{i \in I}$.

Voici maintenant la définition de la dépendance faible au sens de Doukhan et Louhichi (1999) [11] pour les champs aléatoires. Cette notion de dépendance a été essentiellement étudiée dans le cas des séries temporelles et a fait l'oeuvre récemment d'un manuscrit [9]. Voici les définitions que nous allons principalement utiliser dans la suite de ce travail.

Definition 1.3 *Soit* $\|(s_1,\dots,s_d)\| = \max\{|s_1|,\dots,|s_d|\}$ *pour* $s_1,\dots,s_d \in \mathbb{Z}$. *Un champ aléatoire* $(X_t)_{t\in\mathbb{Z}^d}$ *à valeurs dans* $E = \mathbb{R}^k$ *est dit faiblement dépendant si pour une suite* $(\varepsilon_X(r))_{r\in\mathbb{N}}$ *de limite* 0

$$|Cov\,(f(X_{s_1},\dots,X_{s_u})\,,\,g\,(X_{t_1},\dots,X_{t_v})| \leqslant \psi(u,v,Lip\,f,Lip\,g)\varepsilon_X(r),$$

où les indices $s_1,\dots,s_u,\ t_1,\dots,t_v \in \mathbb{Z}^d$ *sont tels que* $\|s_k - t_l\| \geq r$ *pour* $1 \leq k \leq u$ *and* $1 \leq l \leq v$. *De plus, les fonctions à valeurs réelles* f, g *définies sur* $\left(\mathbb{R}^k\right)^u$ *et* $\left(\mathbb{R}^k\right)^v$, *satisfont* $\|f\|_\infty$, $\|g\|_\infty \leqslant 1$ *et* $Lip\,f, Lip\,g < \infty$ *où une norme* $\|\cdot\|$ *est donnée sur* $\mathbb{R}^k$ *et,*

$$Lip\,f = \sup_{(x_1,\dots,x_u)\neq(y_1,\dots,y_u)} \frac{|f(x_1,\dots,x_u) - f(y_1,\dots,y_u)|}{\|x_1-y_1\| + \cdots + \|x_u - y_u\|}.$$

Si $\psi(u,v,a,b) = au + vb$, *on parle de* $\eta-$*dépendance et la suite* $\varepsilon_X(r)$ *sera notée* $\eta_X(r)$.
Si $\psi(u,v,a,b) = au + bv + abuv$, *on parle* $\lambda-$*dépendance et la suite* $\varepsilon_X(r)$ *sera notée* $\lambda_X(r)$.
Si $\psi(u,v,a,b) = au+bv+abuv+u+v$, *on parle de* $\omega-$*dépendance et la suite* $\varepsilon_X(r)$ *sera notée* $\omega_X(r)$.

Un exemple de champs aléatoires $\eta-$faiblement dépendants est celui des schémas de Bernoulli spatiaux, c'est-à-dire de la forme :

$$X_t = H\left((\xi_{t-j})_{j\in\mathbb{Z}^d}\right), \quad t \in \mathbb{Z}^d, \tag{1.1}$$

où ξ est un champ i.i.d et H une fonction mesurable. Cette notion de dépendance apparaît dans l'article de Doukhan et Lang (2002) [10].
La notion de $\lambda-$dépendance a été introduite par Doukhan et Wintenberger (2007) ([14]) dans le cadre des séries, et permet de considérer des schémas de Bernoulli (1.1) suffisamment régulier avec un champ ξ associé.
La notion de $\omega-$dépendance plus générale que la précédente permet de considérer en plus des schémas de Bernoulli (1.1) lorsque ξ est un champ fortement mélangeant sous réserve que $\alpha_\xi(r,a,b) \leq (a+b)u_r$, $a, b \in \mathbb{N}^*$ et la suite u vérifie $\lim_{r\to\infty} u_r = 0$.
Les trois mesures de dépendance faibles que nous venons d'introduire sont stables si on considère des fonctionnelles de type Lipschitz du champ. Ainsi si l'innovation ξ est η, λ ou $\omega-$faiblement dépendante alors on peut montrer qu'il en est de même pour le champ (1.1). Il faut bien noter que cette propriété de stabilité fait défaut dans le cadre de l'association ou du mélange fort. En effet par exemple, il est bien connue que dans le cas d'une chaîne de Markov,

$$X_t = F(X_{t-1};\xi_t)$$

les propriétés de mélange fort sont liées à l'existence d'une partie absolument continue pour l'innovation. De plus il existe des exemples de telles chaînes non mélangeantes, par exemple le modèle AR d'Andrews [1] et nous donnons dans le chapitre 2 de cette thèse un autre exemple qui concerne un modèle ARCH. Cette pathologie n'a pas lieu dans le contexte plus général de la dépendance faible.

1.1.1 Champs aléatoires auto-régressifs

Le Chapitre 2 de cette thèse est un travail joint avec Paul Doukhan. Nous montrons l'existence de champs aléatoires strictement stationnaires à valeurs dans $E = \mathbb{R}^k$ de la forme (1.1) et solutions d'équations autorégressives du type :

$$X_t = F\left((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}}; \xi_t\right). \tag{1.2}$$

Le champ aléatoire ξ est supposé stationnaire et à valeurs dans un espace E', le plus souvent $E' = \mathbb{R}^{k'}$ et la fonction F définie sur $E^{(\mathbb{Z}^d\setminus\{0\})} \times E'$ ([1]) Lipshitz contractante ce qui permet d'appliquer le théorème de point fixe de Picard pour prouver l'existence et l'unicité d'une solution de type (1.1) à l'équation (1.2).
Ce type de construction est nouveau et se différencie des champs de Gibbs dont l'existence est prouvée à l'aide de spécifications conditionnelles et avec des arguments de relative compacité. Ici notre construction utilise des conditions de contraction de type Lipshitz. Voici le résultat principal de ce chapitre :

Théorème 1.1 *On suppose que ξ est un champ stationnaire et que :*

(H1) $\|F(0;\xi_0)\|_m < \infty$.

(H2) *Il existe des réels $a_j \geqslant 0$, pour $j \in \mathbb{Z}^d$ et $\alpha > 0$ tels que $\sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j = e^{-\alpha} < 1$, vérifiant pour tout $z, z' \in E^{(\mathbb{Z}^d\setminus\{0\})}$,*

$$\|F(z;\xi_0) - F(z';\xi_0)\| \quad \leq \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j \|z_j - z'_j\|, \quad a.s. \tag{1.3}$$

Alors il existe dans $\mathbb{L}^m$ une unique solution stationnaire à l'équation (1.2). Cette solution s'écrit sous la forme (1.1).

Au sujet de la dépendance faible dans le cas d'innovations i.i.d., nous avons obtenu :

Théorème 1.2 *Supposons ξ i.i.d. ainsi que* **(H1)** *et* **(H2)**. *Alors la solution de l'équation (1.2) obtenue dans le Théorème 1.1 est $\eta-$faiblement dépendante et il existe une constante $C > 0$ telle que*

$$\eta_X(r) \leqslant C \cdot \inf_{p\in\mathbb{N}^*} \Big\{e^{-\alpha\frac{r}{2p}} + \sum_{\|i\|>p} a_i\Big\}. \tag{1.4}$$

[1]Si V est un espace vectoriel et B un ensemble quelconque alors $V^{(B)} \subset V^B$ désigne l'ensemble des suites $v = (v_b)_{b\in B}$ pour lesquelles il existe un sous ensemble fini $B_1 \subset B$ avec $v_b = 0$ pour $b \notin B_1$.

Nous montrons également l'hérédité de la $\eta, \lambda-$dépendance faible : Si le champ aléatoire ξ est $\eta, \lambda-$faiblement dépendant alors le champ X solution de l'équation (1.2) est aussi $\eta, \lambda-$faiblement dépendant lorsque on remplace l'hypothèse **(H2)** par :

(H2') Il existe un sous-ensemble $\Xi \subset E'$ avec $\mathbb{P}(\xi_0 \in \Xi) = 1$, des nombres positifs $(a_j)_{j\in\mathbb{Z}^d}$ tels que $\sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j = e^{-\alpha} < 1$ avec $\alpha > 0$ et une constante $b > 0$ telle que

$$\|F(x;u) - F(x';u')\| \leq \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j \|x_j - x'_j\| + b\left\|u - u'\right\|,$$

pour $x, x' \in E^{(\mathbb{Z}^d\setminus\{0\})}$ et $u, u' \in \Xi$.

Proposition 1.1 *Supposons* **(H1)** *et* **(H2')**.

1) *Si le champ aléatoire ξ est $\eta-$faiblement dépendant, avec des coefficients de dépendance $\eta_\xi(r)$, alors X est $\eta-$faiblement dépendant et il existe un nombre $C > 0$ tel que,*

$$\eta_X(r) \leq C \inf_{p\in\mathbb{N}^*} \Big\{ \sum_{\|j\|>p} a_j + \inf_{n\in\mathbb{N}^*} \Big\{a^n + p^n \eta_\xi\left((r - 2pn) \vee 0\right)\Big\}\Big\}.$$

2) *Si le champ aléatoire ξ est $\lambda-$faiblement dépendant, avec des coefficients de dépendance notés $\lambda_\xi(r)$, alors X est $\lambda-$faiblement dépendant et il existe $C > 0$ tel que,*

$$\lambda_X(r) \leq C \inf_{p\in\mathbb{N}^*} \left\{ \sum_{\|j\|>p} a_j + \inf_{n\in\mathbb{N}^*} \Big\{a^n + p^{2n} \lambda_\xi\left((r - 2pn) \vee 0\right)\Big\}\right\}.$$

Dans le cas où ξ est un champ i.i.d., une notion de causalité dans l'équation (1.2) permet d'alléger les conditions de moment sur l'innovation ξ. En effet dans le cas général de l'équation (1.2), il faut supposer l'innovation ξ bornée sauf lorsqu'elle est additive.

Definition 1.4 *Si $A \subset \mathbb{Z}^d \setminus \{0\}$, soit $c(A)$ le cône convexe de $\mathbb{R}^d$ engendré par A,*

$$c(A) = \left\{ \sum_{i=1}^{k} r_i j_i \Big/ \ (j_1, \ldots, j_k) \in A^k, (r_1, \ldots, r_k) \in \mathbb{R}_+^k,\ k \geq 1 \right\}.$$

1) La partie A est dite causale si $c(A) \cap \big(- c(A)\big) = \{0\}$.
2) Si F est mesurable par rapport à la σ-algèbre $\mathfrak{F}_A \otimes \mathcal{B}(E')$ pour une partie A causale, alors l'équation $X_t = F\left((X_{t-j})_{j\in I}; \xi_t\right)$ est dite A-causale.

Pour une partie causale $A \subset \mathbb{Z}^d$, on notera $\widetilde{A} = c(A) \cap \mathbb{Z}^d$. On retrouve alors la notion de demi-plan asymétriques introduite par Helson et Lowdenslager (1959) [12] pour généraliser la théorie de la prédiction sur $\mathbb{Z}^2$. Pour une équation $A-$causale, un ordre de type lexicographique apparaît alors naturellement dans l'ordonnancement des variables du champ. Les modèles envisagés peuvent être vus comme des généralisations au cadre spatial des séries temporelles de type ARCH. Ce type de champ n'avait été envisagé jusqu'alors que dans le cadre linéaire ([13],[17]). Voici le résultat d'existence et d'unicité pour les équations $A-$causales.

Théorème 1.3 *Soit $X_t = F\left((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_t\right)$ une équation A-causale avec une innovation i.i.d. ξ. Nous supposons* **(H1)** *ainsi que la condition suivante :*
(H3) *Il existe une suite de nombres positifs $(a_j)_{j\in\mathbb{Z}^d}$ telle que $\sum_{j\in A} a_j = e^{-\alpha} < 1$ avec $\alpha > 0$ et*

$$\|F(x;\xi_0) - F(x';\xi_0)\|_m \leq \sum_{j\in A} a_j \|x_j - x'_j\|, \qquad \forall x, x' \in E^{(\mathbb{Z}^d\setminus\{0\})}.$$

Alors il existe dans $\mathbb{L}^m$ une unique solution strictement stationnaire X à l'équation (1.2) telle que pour $t \in \mathbb{Z}^d$, X_t soit mesurable par rapport à la tribu $\sigma\left(\xi_{t-j}/j \in \widetilde{A}\right)$. Cette solution est $\eta-$faiblement dépendante ; de plus la relation (2.3) vaut pour une constante $C > 0$.

Des calculs en terme d'espérance conditionnelle sont aisés pour les modèles solutions d'équations $A-$causales du fait de l'ordonnancement des variables. Toutefois l'existence d'une direction privilégiée est contraire au comportement attendu d'un phénomène spatial. Dans le cas général d'une équation de type (1.2) non causale, les calculs en terme d'espérance conditionnelle ou même de moments croisés semblent difficiles d'accés. Seul le cadre des modèles linéaire SAR (Simultaneous AR) [16] avait été envisagé par le passé mais dans le cadre linéaire il existe une factorisation de ces modèles en un modèle CAR (Conditional AR) obtenue à l'aide de la densité spectrale, ce qui permet d'avoir une interprétation du modèle en terme de prédiction linéaire et d'assurer l'estimation des paramètres. Dans le cadre non linéaire d'une équation de type (1.2), le problème de l'estimation paramétrique ou non paramétrique semble difficile pour le cas non causal.
Concernant la simulation de ces champs aléatoires, un algorithme de simulation approchée est donné, conséquence directe de la méthode du point fixe. Par exemple, dans le cas particulier où $F(x;u) = f(x_{j_1},\ldots,x_{j_k};u)$ pour une fonction f satisfaisant **(H1)** et **(H2)** , alors pour un champ ξ donné, la suite de champs aléatoires $(X^n)_n$ définie par :

$$X_t^1 = f(0;\xi_t), t \in \mathbb{Z}^d, \quad X_t^{n+1} = f\big(X_{t-j_1}^n,\ldots,X_{t-j_k}^n;\xi_t\big), \qquad \text{pour } n \geqslant 0$$

converge presque sûrement vers la solution X. On peut remarquer que ce principe de simulation par une telle méthode récursive a des similitudes avec l'échantillonneur de Gibbs pour la simulation des champs markoviens. Nous reviendrons sur cette dernière méthode lors de la simulation de textures.

1.1.2 La faible dépendance

Le Chapitre 3 de cette thèse est un travail joint avec Paul Doukhan et Nathanaël Mayo. Dans ce chapitre le cadre général de la dépendance faible est donnée aussi bien pour les séries chronologiques que pour les champs aléatoires. Nous rappelons les différentes formes de dépendance ainsi que les exemples associés. A travers l'exemple de l'estimation des coefficients d'une série ARCH non markovienne, nous expliquons comment la dépendance faible peut être utilisée pour montrer la normalité asymptotique de l'estimateur des moindres carrés.

Des résultats relativement importants sont ensuite donnés au sujet de la théorie limite des champs aléatoires λ et $\omega-$faiblement dépendants. Tous les résultats seront énoncés dans un cadre plus général que la stricte stationnarité puisque nous supposerons seulement :

(A1) Il existe $m > 2$ tel que $\sup_{j\in\mathbb{Z}^d} \mathbb{E}\,|X_j|^m < \infty$.

Dans la section 4.1, nous montrons des inégalités de moments de type Marcinkiewicz-Zygmund, $\mathbb{E}\Big|\sum_{j\in U} X_j\Big|^q \le C\,|U|^{q/2}$, où U est un bloc de $\mathbb{Z}^d$ (i.e $U = (a,b] = ((a_1,b_1]\times\cdots\times(a_d,b_d])\cap\mathbb{Z}^d$ avec $a,b\in\mathbb{R}^d, a_1<b_1,\dots,a_d<b_d$), $|U|$ désigne le cardinal d'un ensemble fini U et q est un nombre réel ≥ 2. La preuve adapte les idées de Bulinski et Shashkin [6] pour des inégalités de moments d'ordre $q = 2+\delta$, $\delta\in[0,1]$.

Nous prouvons ensuite un théorème central limite pour l'une ou l'autre des types de dépendance envisagée. A cet effet nous utilisons les hypothèses suivantes.

Si X est un champ aléatoire centré et $(D_n)_n$ une suite de sous-ensembles finis de $\mathbb{Z}^d$, $S_n = \sum_{j\in D_n} X_j$, et$\sigma_n^2 = \mathrm{Var}\,(S_n)$, nous supposons :

- $(\mathbf{A_2})$ Le champ aléatoire X est $\lambda-$ faiblement dépendant avec $\lambda_X(r) = \mathcal{O}\left(r^{-\lambda}\right)$, $\lambda > 2d \vee d(m-1)/(m-2)$.
- $(\mathbf{A_2'})$ Le champ aléatoire X satisfait $\omega_X(r) = \mathcal{O}\left(r^{-\omega}\right)$, $\omega > 3d\vee dm/(m-2)$.
- $(\mathbf{A_3})$ $\liminf_{n\to\infty} \sigma_n^2/|D_n| > 0$.

Théorème 1.4 *Si* $(\mathbf{A_1})$, $(\mathbf{A_2})$ *ou* $(\mathbf{A_2'})$ *et* $(\mathbf{A_3})$ *sont vérifiées alors* $\sigma_n^{-1}S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Les résultats précédents (inégalité de moments et tlc) sont enfin utilisés pour déduire un principe d'invariance faible. Nous notons pour $t\in\mathbb{Z}^d$, $S_n(t) = \sum_{j\in(0,nt]} X_j$ avec la convention $S_n(t) = 0$ si $\wedge_{1\leqslant i\leqslant d}\, t_i = 0$.

Sous la $\omega-$dépendance nous supposons :

$(\mathbf{A_2''})$ $\omega_X(r) = \mathcal{O}\left(r^{-\omega}\right)$, $\omega > 3d\vee d\frac{4m-5}{m-2}$.

Le résultat suivant est obtenu :

Théorème 1.5 *Soit* X *est un champ aléatoire centré et stationnaire au second ordre (stationnarité* *faible). Supposons* $(\mathbf{A_1})$, $(\mathbf{A_2})$ *ou* $(\mathbf{A_2''})$, *alors :*

$$n^{-d/2}S_n(t) \overset{\mathcal{D}\left([0,1]^d\right)}{\to} \sigma W(t), \qquad \sigma^2 = \sum_{j\in\mathbb{Z}^d} \mathbb{E}X_0X_j \geqslant .0,$$

où W *est un mouvement Brownien à* d *paramètres.*

L'inégalité de moments prouvée dans ce chapitre sera utilisée dans le chapitre suivant en vue de la construction d'une approximation forte pour les sommes partielles de schémas de Bernoulli à innovations i.i.d.

1.1.3 Le principe d'invariance fort

Le Chapitre 4 étudie un Principe d'Invariance Fort (PIF) pour des schémas de Bernoulli spatiaux (1.1) dans le cas d'innovations i.i.d. Ce type d'approximations concerne l'approximation presque sûre des sommes partielles d'un champ aléatoire par une mouvement brownien multiparamétré.
Plus précisément, soit X indexé par $\mathbb{Z}^d$, de la forme (1.1) et à valeurs dans $\mathbb{R}^K$, $K \geq 1$. For $t \in \mathbb{Z}^d$ and $l \in \mathbb{N}^*$, on définit les $\sigma-$algèbres

$$\mathcal{F}_{t,l} = \sigma\left(\xi_{t-j} / \left\|j\right\|_\infty < l\right),$$

ainsi que $X_{t,l} = \mathbb{E}_{\mathcal{F}_{t,l}}X_t$. Si l est pair on pose $X_{t,l/2} = X_{t,(l+1)/2}$. Pour $l = 0$, on pose $X_{t,l} = X_t$, $\forall t \in \mathbb{Z}^d$.
La mesure de dépendance qui sera utilisé pour le champ X est donnée par les coefficients :

$$p(l) = \left\|X_0 - X_{0,l}\right\|_2, \quad l \in \mathbb{N}^*. \tag{1.5}$$

Nous aurons besoin des hypothèses suivantes :

(A1) $X = (X_t)_{t\in\mathbb{Z}^d}$ est un schéma de Bernoulli (1.1) à valeurs dans $\mathbb{R}^K$, $K \in \mathbb{N}^*$ où $(\xi_t)_{t\in\mathbb{Z}^d}$ est un champ aléatoire i.i.d. De plus il existe $h > 2$ et $\widetilde{C} > 0$ tel que $\mathbb{E}\left\|X_0\right\|^h < \infty$ et pour $r \in \mathbb{N}$, $p(r) \leq \widetilde{C}(r+1)^{-\eta}$ où $\eta > 2d\,\frac{h-1}{h-2}$.

(A2) $\Gamma = \sum_{j\in\mathbb{Z}^d} \Gamma\left(X_0, X_j\right)$ est une matrice définie positive.

La suite de coefficients $(p(l))_{l\in\mathbb{N}^*}$ mesure la qualité d'approximation du schéma de Bernoulli X par un champ aléatoire $m-$dépendant. La convergence de cette suite vers 0 entraîne automatiquement la propriété de $\eta-$faible dépendance.

Le résultat d'approximation utilise les ensembles G_τ, $\tau \in (0,1)$ définis par :

$$G_\tau = \bigcap_{s=1}^{d} \Big\{ j = (j_1, \ldots, j_d) \in \mathbb{N}^d,\, j_s \geq \big(\prod_{s'\neq s} j_{s'}\big)^\tau \Big\}.$$

Nous avons alors obtenu le :

Théorème 1.6 *Sous les hypothèses* **(A1)** *et* **(A2)** *et pour* $\tau \in (0,1)$*, il existe* $\varepsilon > 0$ *et* X *peut être redéfini sans changer sa distribution sur un espace de probabilité plus riche avec un mouvement Brownien à* d *paramètres* $W = \{W_t, t \in [0,\infty)^d\}$ *tel que :*

$$S_N - W_N = O\left([N]^{1/2-\varepsilon}\right), \quad N \in G_\tau \cap \mathbb{Z}^d. \tag{1.6}$$

Ici $S_N = \sum_{j\in(0,N]} X_j$ *et* $[N] = N_1 \cdots N_d$.

Le premier résultat d'approximation forte pour les champs aléatoires a été établi par Berkes et Morrow (1981) [4] dans le cadre des champs fortement mélangeants. Ces auteurs montrent qu'une approximation du type (1.6) ne peut avoir lieu lorsque N est trop proche des axes et ceci conduit à l'introduction de parties "équilibrées" pour l'asymptotique (d'où l'introduction des ensembles G_τ). De la même manière, d'autres PIF pour les champs aléatoires ont été récemment établie pour les champs associés par Balan (2005) [3] et leurs extensions par Bulinski et Shaskin (2005) [6]. Contrairement à ces deux derniers cas, notre preuve ne nécessitent que des décroissances de type polynomiale pour les covariances car nous tirons le bénéfice d'une approximation par des champs aléatoires $m-$dépendants pour pouvoir générer les variables gaussiennes. Récemment, cet avantage a été utilisé dans le cas des séries de type ARCH dans les travaux de Aue et al. (2006) [2] et Liu et Lin (2008)[18]. Ces derniers auteurs obtiennent des résultats de vitesse optimaux dans le sens où ils retrouvent les vitesses du cas i.i.d. Nous n'avons pas investi le problème de l'optimalité dans notre cas (ici déterminé par la valeur de ε) ; le contrôle des moments utilisent les résultats généraux des champs $\eta-$faiblement dépendants alors que dans le cas des séries, l'utilisation de décompositions en différences de martingales rend ce contrôle plus efficace. Il serait alors intéressant d'investir une meilleure inégalité de moments que celle que nous avons utilisée, pour le cas des schémas de Bernoulli à innovations i.i.d.

1.2 Quelques problèmes d'estimation paramétriques et non paramétriques

1.2.1 Simulation de textures par champs de Markov

Dans le Chapitre 5, nous nous intéressons le problème du rééchantillonnage des champs aléatoires et son utilisation dans le cadre de la synthèse de textures par champs markoviens. On appelle texture un motif graphique non trivial et la simulation de textures va consister en la construction d'une image reproduisant la texture initiale mais sur une surface de taille supérieure. La Figure 1.1 ci-dessous donne un exemple d'une telle simulation.

Notre étude est guidée et motivée par la généralité statistique d'une méthode de simulation. A cet effet nous étudions en particulier un algorithme non causal extension du bootstrap de Bickel et Levina (2006) [5] énoncé dans le cadre des champs unilatéraux et dont le but était d'expliquer les bons

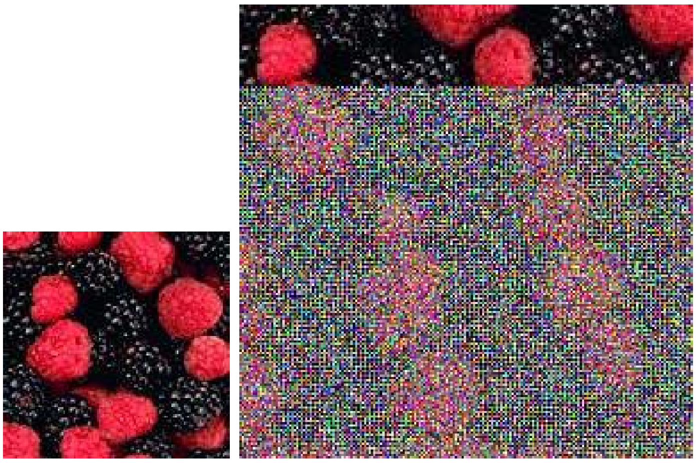


FIG. 1.1: Synthèse d'une texture à partie d'un échantillon.

résultats de l'algorithme de Efros et Leung [16]. Il est d'usage pour les textures de considérer les valeurs des pixels observées comme une réalisation d'un champ markovien stationnaire. L'objectif est alors d'estimer ce champ et d'en simuler une trajectoire approchée pour reproduire une texture similaire mais de taille plus importante. Les techniques qui ont eu le plus de succès comme l'algorithme de Efros et Leung sont non paramétriques. Cependant ces types d'algorithmes n'ont le plus souvent pas de réelles justifications théoriques. Dans cette direction, Bickel et Levina proposent un bootstrap formel consistant pour expliquer ces algorithmes. Nous rappelons tout d'abord le principe de rééchantillonnage formulé par Bickel et Levina.
Pour $o \in \mathbb{N}^*$, $t = (t_1, t_2) \in \mathbb{N}^* \times \mathbb{N}^*$ et $s \in \mathbb{N}^* \times \mathbb{N}^*$, on définit les ensembles :

$$\begin{aligned} U_t &= \big\{u \neq t \in \mathbb{N}^* \times \mathbb{N}^*;\ \max(1, t_1 - o) \leq u_1 \leq t_1,\ \max(1, t_2 - o) \leq u_2 \leq t_2\big\}; \\ U_t(s) &= U_t - \{t\} + \{s\}; \\ W_t &= \{1, \dots, t_1\} \times \{1, \dots, t_2\} \setminus \{t\}. \end{aligned}$$

Bickel et Levina supposent que le champ aléatoire qui définit la texture vérifie la condition suivante :

Definition 1.5 *Un champ $X = \{X_t,\, t \in \mathbb{N}^* \times \mathbb{N}^*\}$ est unilatéral si il existe $o \in \mathbb{N}^*$ tel que pour $t \in \mathbb{N}^* \times \mathbb{N}^*$,*

$$\mathbb{P}\left(X_t / X_{W_t}\right) = \mathbb{P}\left(X_t / X_{U_t}\right). \tag{1.7}$$

Soit maintenant X le champ aléatoire qui modélise une texture, observé sur $\{1, \dots, T_1\} \times \{1 \dots, T_2\}$ *i.e.*

$$\left(X_t,\, t \in \{1, \dots, T_1\} \times \{1 \dots, T_2\}\right)$$

est connue. On considère ensuite un noyau $W : \mathbb{R}^\ell \to \mathbb{R}$. De plus pour une largeur de fenêtre $b > 0$, on définit :

$$W_b(y) = b^{-\ell} W(y/b) \text{ for all } y \in \mathbb{R}^\ell.$$

Dans la suite, on note $X^* = \{X_t^*,\, t \in \mathbb{N}^* \times \mathbb{N}^*\}$ la texture générée à l'aide de $(X_t,\, t \in \{1, \dots, T_1\} \times \{1 \dots, T_2\})$. Voici l'algorithme :

1. On choisit $\{X_t^* : 1 \leq t_1 \leq o+1, 1 \leq t_2 \leq o+1\}$ uniformément dans l'ensemble de carrés de taille $(o+1) \times (o+1)$ de $(X_t,\, t \in \{1, \dots, T_1\} \times \{1 \dots, T_2\})$.
2. Supposons que pour $(u, v) \in \mathbb{N}^* \times \mathbb{N}^*$, X_t^* a été généré pour $t \in \{1, \dots, u-1\} \times \{1, \dots, v\} \cup \{u\} \times \{1, \dots, v-1\}$, c'est à dire, $u-1$ lignes sont totalement remplies, et la ligne u est remplie jusqu'à la colonne v. Pour générer la valeur $X_t^* = X_{(u,v)}^*$, soit N_t une variable aléatoire discrète de loi :
$$\mathbb{P}(N_t = s) = \frac{1}{Z} W_b\big(X_{U_t}^* - X_{U_t(s)}\big)$$
pour $s \in \mathbb{N}^* \times \mathbb{N}^*$ such that $U_t(s) \subset \{1, \dots, T_1\} \times \{1 \dots, T_2\}$ et où $Z = \sum_s W_b(Y_t^* - Y_t(s))$ est une constante de normalisation. La somme précédente est indexée par l'ensemble des sites s dont le voisinage considéré est inclus dans la texture observée.
3. On génère N_t et on pose $X_t^* = X_{(u,v)}^* = X_{N_t}$.

La base de cette méthode est l'estimation non paramétrique de la fonction de distribution conditionnelle ponctuelle du champ. En effet, si A est une partie finie de $\mathbb{Z}^2$, et $Y_t = X_{t+A}$, $t \in \mathbb{Z}^2$, Bickel et Levina estime la distribution $F_{X/Y}(x/y) = \mathbb{P}(X \leq x/Y = y)$ par :

$$F_T(dx/y) = \frac{\sum_{s \in I_T} W_b(y - Y_s)\delta_{X_s}(dx)}{\sum_{s \in I_T} W_b(y - Y_s)},$$

où I_T est l'ensemble des s tel que (X_s, Y_s) soit observé. L'étape 2 de l'algorithme ci-dessus indique que $F_{X_t^*/Y_t^*}(dx/y) = F_T(dx/y)$ pour $A = U_t$.

Les hypothèses introduites par Bickel et Levina seront aussi nos hypothèses de travail.

(A1) Le champ aléatoire X est strictement stationnaire et α-mixing, *i.e.* si pour $k, u, v \in \mathbb{N}^*$,

$$\alpha_X(k, u, v) = \sup_{E,F \in \mathbb{Z}^2, d(E,F)=k, |E|=u, |F|=v} \big\{|P(AB) - P(A)P(B)|,\, A \in \sigma(X_E), B \in \sigma(X_F)\big\}$$

sont les coefficients de mélange tels que il existe $\varepsilon > 0$, $\tau > 2$ qui satisfait pour tous les entiers $u, v \geq 2$, $u + v \leq c$, où c est le plus petit entier pair tel que $c \geq \tau$,

$$\sum_{k=1}^{\infty} (k+1)^{2(c-u+1)-1} \alpha_X(k, u, v)^{\varepsilon/(c+\varepsilon)} < \infty.$$

(A2) X_t a un support compact $S \subset \mathbb{R}$.

(A3) $F_{X,Y} = \mathbb{P}(X_t \leq \cdot, Y_t \leq \cdot)$, $F_{X/Y}$ et $F_Y = \mathbb{P}(Y_t \leq \cdot)$ ont des densités bornées, continues et strictement positives (notées respectivement $f_{X,Y}$, $f_{X/Y}$ et f_Y) par rapport à la mesure de Lebesgue. De plus, il existe $L > 0$ tel que pour $y, y' \in S^A$, $x \in S$,

$$\left| \int_{-\infty}^{x} f_{X,Y}(z,y)dz - \int_{-\infty}^{x} f_{X,Y}(z,y')dz \right| \leq L \left\| y - y' \right\|.$$

(A4) Si l est la dimension de Y_t, le noyau $W : \mathbb{R}^{\ell} \to (0,\infty)$ est borné, symétrique et Lipshitz. De plus,

$$\int u\, W(u)du = 0 \quad \text{and} \quad \int \|u\|\, W(u)du < \infty.$$

On définit $W_{b_T}(u) = b_T^{-l} W(u/b_T)$, où $b_T = O([T]^{-\delta})$, with $\delta > 0$.

Bickel et Levina ont montré la convergence uniforme presque sûre de l'estimateur F_T vers $F_{X/Y}$. Nous donnons en plus une indication quant à la vitesse de convergence presque sûre. Notons que cette convergence est indépendante d'une quelconque hypothèse markovienne sur X

Théorème 1.7 *Si X est un champ aléatoire qui satisfait* $(\mathbf{A1-4})$, *alors pour toute partie A finie de $\mathbb{Z}^2$:*

$$\sup_{(x,y)\in S\times S^A} \left| F_T(x/y) - F_{X/Y}(x/y) \right| = O\left([T]^{-\gamma}\right) \quad a.s$$

où $0 < \gamma < \dfrac{\tau - 2}{2(v+1)(\tau+v+2)}$ *et* $b = b_T = O([T]^{-\delta})$ *avec* $\delta = \frac{\tau-2}{2(v+1)(\tau+v+2)}$.

Bickel et Levina prouve ensuite la consistance de leur bootstrap ; sous l'hypothèse d'unilatéralité les lois jointes sont reconstructibles à partir des lois conditionnelles.
Notre objectif a été ensuite de voir si les distributions conditionnelles F_T pouvaient être utilisées dans le but d'une simulation non causale. A partir de maintenant, nous posons

$$A = \mathcal{N}_o = \{j \in \mathbb{Z}^2 / 0 < \|j\|_\infty \leq o\}.$$

Les champs unilatéraux (5.1) sont des cas très particuliers des champs de Markov qui vérifient :

$$\mathbb{P}(X_t/X_s, s \neq t) = \mathbb{P}\left(X_t/X_{t+\mathcal{N}_o}\right), \quad t \in \mathbb{Z}^2. \tag{1.8}$$

L'intérêt de notre recherche est bien sûr d'avoir un résultat de consistance pour un champ aléatoire markovien stationnaire en général satisfaisant (1.8), sans l'hypothèse d'unilatéralité supposée par Bickel et Levina. Les problèmes mathématiques qui interviennent pour cette généralisation sont plus difficiles. En effet, dans le cas non causal, les lois jointes du champ ne sont pas directement reconstructibles à partir des lois conditionnelles. De plus il n'existe pas en général de champ aléatoire admettant $F_T(dx/y)$ pour distributions conditionnelles ponctuelles (c'est-à-dire les lois d'un site sachant les autres). Nous référons à [16] pour une explication de ce problème et pour la définition de champ de Markov qui se fait à l'aide d'un potentiel d'interactions. La définition d'un champ de Markov nécessite d'avoir toute une famille de lois conditionnelles sur toutes les parties finies de l'espace sachant leur complémentaire.

Nous avons alors considéré une technique de rééchantillonnage à partir des méthodes usuelles de simulation pour les champs aléatoires, comme l'échantillonneur de Gibbs dont nous rappelons le principe. Soit Z est un champ de Markov sur $\mathcal{X}$, et R est un rectangle de $\mathbb{Z}^2$ avec :

$$\partial R = (R + \mathcal{N}_o) \setminus R.$$

Fixons pour $i \in \partial R$, des valeurs arbitraires $x_i \in S$. Pour simuler à l'aide de l'échantillonneur de Gibbs une réalisation approchée de $\mathbb{P}\left(Z_R / Z_{\partial R} = x_{\partial R}\right)$, on initialise l'échantillonneur en choisissant une valeur $z(0) \in S^R$. Une suite d'images $z(i) \in S^R$, $i \in \mathbb{N}^*$, est alors définie en visitant une infinité de fois chaque site $s \in R$ et en remplaçant à chaque fois la valeur du pixel au site s par une réalisation de la distribution conditionnelle $\mathbb{P}\left(Z_s / Z_{s+\mathcal{N}_o} = y\right)$, où y représente les valeurs des pixels aux sites voisins de s dans $R \cup \partial R$. On peut alors montrer que :

$$\lim_{i \to \infty} z(i) = \mathbb{P}\left(Z_R / Z_{\partial R} = x_{\partial R}\right) \quad \text{en loi.}$$

Nous avons montré que si on utilise les distributions F_T, avec un schéma de visites de type lexicographique dans l'échantillonneur de Gibbs, on obtient une mesure limite ν_T sur S^R et que :

$$\lim_{T \to \infty} \nu_T = \mathbb{P}\left(X_R / X_{\partial R} = x_{\partial R}\right), \quad \text{a.s,}$$

la limite précédente étant la limite en loi. Il resterait en fait à étudier l'asymptotique véritable de ce bootstrap, en supposant que pour T fixé on fait tourner l'échantillonneur de Gibbs à l'aide des lois F_T sur un rectangle R_T tel que $R_T \nearrow \mathbb{Z}^2$, puisque une simulation de textures se fait sur un rectangle plus gros que le rectangle I_T des observations. Nous n'avons pu trouver de preuve dans ce cas.

Pour appliquer concrètement ce bootstrap non causal à la synthèse de texture, d'autres techniques de simulations sont nécessaires. En effet les longs temps de relaxation stochastique et la nécessité de prendre en compte les caractéristiques de la texture à différentes résolutions demandent un algorithme adapté emprunté à Paget et Longstaff (1998) [19]. Ces auteurs avaient étudié une méthode similaire à la notre mais dont la validation reposait uniquement sur des simulations. Nous avons donné des exemples convaincants de cette méthode pour la simulation de texture, en utilisant les grilles multirésolutions et la fonction de température introduites par Paget et Longstaff. En effet les méthodes de type causales pixel par pixel marchent très bien mais certaines simulations montrent une tendance à "casser" lorsque l'on s'éloigne du germe associé à l'initialisation de ces algorithmes. Il faut tout de même noter que les méthodes récentes de simulations de textures utilisent directement des blocs au lieu de générer les pixels un par un. Dans ce cas les problèmes de type "cassure" des méthodes causales peuvent être éviter et les temps de calcul sont relativement courts. Cependant ces méthodes présentent d'autres défauts.

1.2.2 Un modèle de type bilinéaire pour des séries à valeurs entières

Au Chapitre 6, dans un travail joint avec Alain Latour, nous définissons un nouveau modèle autorégressif à valeurs entières. Deux éléments nouveaux sont introduits pour prendre en compte deux nouvelles contraintes pour ce type de modèle. Tout d'abord la nécessité d'avoir un modèle qui prend des valeurs aussi bien positives que négatives. Une telle contrainte est naturelle si on est amené à stationnariser par différences un processus à valeurs entières positives. Ensuite l'observation de certaines séries suggère le besoin d'une hétéroscédasticité plus forte pour pouvoir expliquer des instabilités dans le comportement des données.
Le modèle que nous avons étudié utilise une extension d'un opérateur connu dans la construction des modèles à valeurs entières positives : l'opérateur d'amincissement (voir [21]).

Definition 1.6 (Opérateur d'amincissement signé) *Soit $Y = (Y_i)_{i\in\mathbb{N}}$ une suite i.i.d. de variables aléatoires à valeurs entières d'espérance α et indépendantes d'une variable à valeurs entières X. L'opérateur $\alpha\circ$ est défini par :*

$$\alpha \circ X = \begin{cases} sign(X)\sum_{i=1}^{|X|} Y_i, & si \quad X \neq 0; \\ 0, & sinon. \end{cases}$$

Dans la définition précédente, la notation $sign(x)$ pour un réel x vaut pour le signe de x (1 si $x > 0$, -1 sinon). Le modèle que nous étudions dans ce chapitre est défini par :

$$X_t = \sum_{j=1}^{\infty} \alpha_j \circ X_{t-j} + \varepsilon_t \sum_{j=1}^{\infty} \beta_j \circ X_{t-j} + \eta_t. \tag{1.9}$$

On suppose la suite $\mathbb{E}\varepsilon_t = 0$ et pour $t \in \mathbb{Z}$, soit :

$$\xi_t = \left((Y_{t,i}^{(j)})_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*}, (\widetilde{Y}_{t,i}^{(j)})_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*}, \varepsilon_t, \eta_t \right).$$

On suppose la suite $(\xi_t)_{t\in\mathbb{Z}}$ i.i.d. et pour tout $t \in \mathbb{Z}$, ε_t indépendante de la suite $\left(\widetilde{Y}_{t,i}^{(j)}\right)_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*}$.

On peut observer l'analogie avec les modèles de type ARCH introduits dans le cas d'un espace d'états continu, spécialement avec le modèle bilinéaire introduit dans [22].

Théorème 1.8 *Supposons que pour un entier $m \geq 1$,*

$$a = \sum_{j=1}^{\infty} \left\| Y_1^{(j)} \right\|_1 + \|\varepsilon_0\|_m \left\| \widetilde{Y}_1^{(j)} \right\|_1 < 1, \qquad \sum_{j=1}^{\infty} \left\| Y_1^{(j)} \right\|_m + \left\| \widetilde{Y}_1^{(j)} \right\|_m + \|\eta_0\|_m < \infty, \tag{1.10}$$

alors il existe une unique solution stationnaire à l'équation (1.9) telle que :
- $\mathbb{E}\,\|X_0\|^m < \infty$.
- $X_t \in \sigma\left(\xi_t, \xi_{t-1}, \ldots\right)$, $t \in \mathbb{Z}$.

La méthode utilisée pour prouver le théorème 1.8 utilise des résultats du Chapitre 1 énoncés pour le cadre général des champs aléatoires causaux. Nous avons aussi utilisé une inégalité de moments pour l'opérateur d'amincissement (Lemme 3.1, point 4) déjà utilisée par Drost et al. [14].

Soit $p \in \mathbb{N}^*$ fixé. Le modèle paramétrique que nous avons considéré est le suivant :

$$X_t = \sum_{j=1}^{p} \alpha_j \circ X_{t-j} + \varepsilon_t \sum_{j=1}^{p} \beta_j \circ X_{t-j} + \eta_t. \tag{1.11}$$

Lorsque $\varepsilon_t = 0$, $t \in \mathbb{Z}$, on retrouve (modulo l'extension sur $\mathbb{Z}$ de l'opérateur d'amincissement) un modèle connu, le Ginar(p) (voir [15] et [26]), qui est en fait l'équivalent, à valeurs entières, du modèle AR du point de vue de la structure de covariance. Nous avons procédé à une estimation des coefficients du modèle (1.11) à l'aide de l'estimateur du quasi-maximum de vraisemblance gaussien (QMLE, abréviation issue de l'anglais). Voici le problème d'estimation que nous avons envisagé et les hypothèses que nous avons utilisé.
Pour $(t, j) \in \mathbb{Z} \times \{1, \ldots, p\}$, on définit les $\sigma-$algebres :

$$\mathcal{F}_t = \sigma(X_{t-k} : k \in \mathbb{N}), \quad \mathcal{G}_{t,j} = \sigma(Y_{t,i}^{(j)} : i \in \mathbb{N}^*), \quad \text{and} \quad \widetilde{\mathcal{G}}_{t,j} = \sigma(\widetilde{Y}_{t,i}^{(j)} : i \in \mathbb{N}^*),$$

ainsi que les hypothèses suivantes :

1. $\forall t \in \mathbb{Z}$, les σ-algebres $\mathcal{G}_{t,1}, \ldots, \mathcal{G}_{t,p}$ (resp. $\widetilde{\mathcal{G}}_{t,1}, \ldots, \widetilde{\mathcal{G}}_{t,p}$) sont indépendantes.
2. Pour $t \in \mathbb{Z}$, les σ-algebres $\sigma(\varepsilon_t)$, $\sigma(\eta_t)$ et $(\vee_{1 \leq j \leq p} \mathcal{G}_{t,j}) \vee (\vee_{1 \leq j \leq p} \widetilde{\mathcal{G}}_{t,j})$ sont mutuellement indépendantes.

Remark Pour un entier $d \geq 1$, soit Θ a sous ensemble de $\mathbb{R}^d$ et $\theta_0 \in \Theta$.

Pour $1 \leq j \leq p$, on considère les fonctions $b_j, c_j, w_j, \mu, \nu : \Theta \to \mathbb{R}$ telles que :
i) $b_j(\theta_0) = \alpha_j$ and $c_j(\theta_0) = \beta_j$. Pour assurer l'identification, on suppose qu'il existe $j_0 \in \{1, \ldots, p\}$ tel que $\beta_{j_0} > 0$ et que la fonction c_{j_0} est positive sur Θ.
ii) $w_j(\theta_0) = \mathrm{Var}\,(Y^{(j)}) + \sigma^2 \times \mathrm{Var}\,(\widetilde{Y}^{(j)})$, $\sigma^2 = \mathrm{Var}\,\varepsilon_0$.

iii) $\mu(\theta_0) = \mathbb{E}\eta_0$ et $\nu(\theta_0) = \text{Var}(\eta_0)$.
Nous utilisons aussi les hypothèses suivantes :

H1) Θ est un compact de $\mathbb{R}^d$.

H2) Condition (1.10) est vérifiée pour $m = 2$.

H3) Le support de la distribution de η_0 contient au moins 5 points distincts si $\text{Var}(\varepsilon_0) \neq 0$ et 3, sinon.

H4) La fonction ν satisfait : $h = \inf_{\theta\in\Theta} \nu(\theta) > 0$.

H5) La fonction $f : \Theta \to \mathbb{R}^{3p+2}$ définie par :

$$f(\theta) = \Big((b_j(\theta), c_j(\theta), w_j(\theta))_{1\leq j\leq p}, \mu(\theta), \nu(\theta) \Big)$$

est injective et continue sur Θ.

Pour $(t, \theta) \in \mathbb{Z} \times \Theta$, soit

$$m_t(\theta) = \mu(\theta) + \sum_{j=1}^{p} b_j(\theta) X_{t-j}$$

et

$$V_t(\theta) = \sigma^2 \left(\sum_{j=1}^{p} c_j(\theta) X_{t-j} \right)^2 + \sum_{j=1}^{p} w_j(\theta) |X_{t-j}| + \nu(\theta).$$

On a alors :

$$\mathbb{E}(X_t/\mathcal{F}_{t-1}) = m_t(\theta_0), \quad \text{Var}(X_t/\mathcal{F}_{t-1}) = V_t(\theta_0).$$

On peut alors observer que la moyenne conditionnelle est la même que pour le processus GINAR(p). En revanche, la variance conditionnelle diffère du GINAR(p) par une partie polynomiale de degré 2.

Concernant l'estimation des paramètres, nous utilisons le QMLE. Voici le principe :
Supposons que $X_0, \ldots, X_{-p+1}$ sont observées. On définit le contraste :

$$\begin{aligned} q_t(\theta) &= \frac{(X_t - m_t(\theta))^2}{V_t(\theta)} + \ln V_t(\theta), \qquad t \in \mathbb{Z}; \\ Q_T(\theta) &= \frac{1}{T} \sum_{t=1}^{T} q_t(\theta); \\ Q(\theta) &= \mathbb{E}\left(\left(\frac{(X_0 - m_0(\theta))^2}{V_0(\theta)} + \ln V_0(\theta) \right) \right); \\ \hat{\theta}_T &= \arg\min_{\theta\in\Theta} Q_T(\theta). \end{aligned}$$

Nous avons alors prouvé le résultat de consistance suivant :

Théorème 1.9 *Sous les hypothèses* **H1)** *à* **H5)***, on a :* $\hat{\theta}_T \to_{a.s.} \theta_0$.

Dans la suite, si g est une fonction, $g : \Theta \mapsto \mathbb{R}$, on note ∇g son gradient et $\nabla^2 g$ sa matrice hessienne.

Nous avons également prouvé la normalité asymptotique de notre estimateur, avec des hypothèses supplémentaires :

H7) La condition (1.10) est satisfaite pour $m = 4$.

H8) La fonction f est deux fois différentiable sur Θ et $rank \nabla f(\theta_0) = d$. De plus, $\inf_{\theta \in \Theta} w_j(\theta) > 0, \forall j = 1, \ldots, p$.

H9) θ_0 est un point intérieur à Θ.

Théorème 1.10 *Sous les hypothèses* **H1)***, ...,* **H9)** *:*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \to_{T \to \infty} \mathcal{N}(0, F_0^{-1} G_0 F_0^{-1}) \quad \textit{en loi,}$$

où

$$\begin{aligned} F_0 &= \mathbb{E}\left(\nabla^2 q_0(\theta_0)\right) \\ &= \mathbb{E}\left(V_0(\theta_0)^{-2} \nabla V_0(\theta_0) \nabla V_0(\theta_0)'\right) + 2\mathbb{E}\left(V_0(\theta_0)^{-1} \nabla m_0(\theta_0) \nabla m_0(\theta_0)'\right) \end{aligned}$$

et

$$\begin{aligned} G_0 = {} & Var\,(\nabla q_0(\theta_0)) \\ & = \mathbb{E}\left(V_0(\theta_0)^{-4}(X_0 - m_0(\theta_0))^4 \nabla V_0(\theta_0) \nabla V_0(\theta_0)'\right) \\ & \quad - \mathbb{E}\left(V_0(\theta_0)^{-2} \nabla V_0(\theta_0) \nabla V_0(\theta_0)'\right) + 4\mathbb{E}\left(V_0(\theta_0)^{-1} \nabla m_0(\theta_0) \nabla m_0(\theta_0)'\right) \\ & \quad + \mathbb{E}\left(V_0(\theta_0)^{-3}(X_0 - m_0(\theta_0))^3 \nabla V_0(\theta_0) \nabla m_0(\theta_0)'\right) \\ & \quad + \mathbb{E}\left(V_0(\theta_0)^{-3}(X_0 - m_0(\theta_0))^3 \nabla m_0(\theta_0) \nabla V_0(\theta_0)'\right) \end{aligned}$$

Enfin nous avons déduit de notre approche une généralisation dans l'utilisation paramétrique du modèle GINAR(p). En effet, l'estimation de ce modèle dans [15] ou [26] utilise les moindres carrés conditionnels pour estimer le paramètre des opérateurs d'amincissements (en fait les moyennes α_j dans (1.11)). Grâce au QMLE, on peut en fait estimer 2 paramètres de ces opérateurs. Nous référons à la section 5 de ce Chapitre pour de plus amples détails.
Une perspective pour ce travail sera de confronter ce modèle avec des données réelles, comme par exemple la série représentée dans l'introduction du Chapitre 6.

1.2.3 QMLE lissé et estimation des modèles LARCH

Dans le Chapitre 7, une nouvelle méthode d'estimation est considérée. Les modèles de type LARCH possèdent un problème majeur en vue de leur identification, du fait que leur variance asymptotique ne soit pas bornée inférieurement. Ces modèles ont été introduits dans le cadre de la longue portée (voir [20], [21] et [22]). Dans un contexte de courte portée, le modèle autorégressif que nous envisageons est du type

$$\begin{aligned} Y_t &= b_{0,1}Y_{t-1} + \cdots + b_{0,q}Y_{t-q} + X_t, && (1.12)\\ X_t &= \xi_t \left(a_{0,0} + \sum_{j=1}^{p} a_{0,j}X_{t-j} \right), \qquad t \in \mathbb{Z} && (1.13) \end{aligned}$$

avec ξ une suite i.i.d. telle que $\mathbb{E}\xi_0 = 0$, $\mathbb{E}\xi_0^2 = 1$. L'existence d'une solution stationnaire au modèle (1.12) a été étudié dans [18].
Posons

$$\theta_0 = (b_{0,1}, \ldots, b_{0,q}, a_{0,0}, \ldots, a_{0,p}),$$

Notre but ici est d'estimer ce paramètre θ_0. Un travail récent de Francq et Zakoïan [19] étudie ce même problème à l'aide d'une méthode par méthode des moindres carrés pondérés. Ces auteurs montrent également que le QMLE Gaussien déjà présenté au Chapitre 5 est inconsistent en général pour le modèle (1.12). Nous avons proposé une méthode différente à l'aide d'un contraste pénalisé, qui a été aussi introduite récemment par Beran et Shützner [5] dans le cadre du modèle LARCH mais pour l'estimation de paramètres dans le cadre de la longue portée.
Pour $\theta = (b_1, \ldots, b_q, a_0, \ldots, a_p) \in \mathbb{R}^{p+q+1}$ et $t \in \mathbb{Z}$, on pose :

$$\begin{aligned} m_t(\theta) &= \sum_{j=1}^{q} b_j Y_{t-j},\\ V_t(\theta) &= \sigma_t^2(\theta) = \left(a_0 + \sum_{j=1}^{p} a_j \left(Y_{t-j} - m_{t-j}(\theta) \right) \right)^2. \end{aligned}$$

Si $\mathcal{F}_t = \sigma\left(Y_{t-1}, Y_{t-2}, \ldots\right)$ on a pour $t \in \mathbb{Z}$:

$$m_t(\theta_0) = \mathbb{E}\left(Y_t/\mathcal{F}_{t-1}\right), \qquad V_t(\theta_0) = \mathrm{Var}\left(Y_t/\mathcal{F}_{t-1}\right).$$

Par stationnarité, on peut toujours supposer que les données $Y_n, Y_{n-1}, \ldots, Y_{-(p+q)+1}$ sont observées. On définit alors pour un paramètre de lissage $h \geq 0$:

$$\hat{\theta}_{h,n} = \arg\min_{\theta \in \Theta} Q_{h,n}(\theta),$$

$$Q_{h,n}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \frac{(Y_t - m_t(\theta))^2 + h}{V_t(\theta) + h} + \log\left(V_t(\theta) + h\right). \tag{1.14}$$

Pour $h = 0$, on reconnait le QMLE qui ne semble pas applicable ici. Nous avons montré que pour tout $h > 0$, sous certaines conditions, notre estimateur $\hat{\theta}_{h,n}$ était consistant et asymptotiquement normal (Théorèmes 1 et 2 de ce Chapitre).

Nous avons ensuite étudié le comportement de la variance asymptotique de notre estimateur en fonction de h. Le Lemme 3 de ce chapitre discute les cas ou la plus petite variance est atteinte lorsque $h \to 0$ (c'est en particulier le cas pour un modèle LARCH (1.13)). Cette variance limite peut être dégénérée.

Des exemples de simulations dans ce chapitre suggèrent la nécessité de trouver un équilibre pour la valeur de h. Si on diminue h, l'erreur quadratique semble moins importante sous réserve que la taille n de l'échantillon soit assez importante. Une méthode pour déterminer $h = h_n$ serait alors intéressante à étudier, ainsi qu'une asymptotique adaptée. La confrontation de ce modèle avec des données réelles est cependant envisageable à h fixée et serait intéressante à poursuivre.

Bibliographie

[1] Andrews, D. (1984) *Non strong mixing autoregressive processes.* Journal of Applied Probability 21, 930-934.

[2] Aue, A., Berkes, I., Horváth, L. (2006) *Strong approximation for sums of squares of augmented GARCH sequences.* Bernoulli, 12, pp. 583-608.

[3] Balan, R. (2005) *A strong invariance principle for associated random fields.* The Annals of Probability, Vol. 33, No. 2, 823-840.

[4] Berkes, I., Morrow, G. J. (1981) *Strong invariance principle for mixing random fields.* Z. Wahrsh. Verw. Gebiete 57 15-37.

[5] Beran, J., Schützner, M. (2008) *Estimation of dependence parameters for linear ARCH processes.* Working document.

[6] Bickel, P., Levina, E. (2006) *Texture synthesis and nonparametric resampling of random fields* The Annals of Statistics, Vol. 34, No 4, 1751-1773.

[7] Bulinski, A., Shashkin, A. (2006)*Strong invariance principle for dependent random fields* IMS Lecture Notes–Monograph Series, Vol. 48, 128-143.

[8] Bulinski, A., Shashkin, A. *Limit Theorems for Associated Random Fields and Related Systems.* Advanced Series on Statistical Science and Applied Probability. Vol 10.

[9] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, and S., Prieur, C. (2007) *Weak dependence, examples and applications*, Lecture Notes in Statistics 190.

[10] Doukhan, P. (1994) *Mixing, properties and examples.* Lecture Notes in Statistics 85, Springer Verlag, New-York.

[11] Doukhan, P. and Lang, G. (2002) Rates in the empirical central limit theorem for stationary weakly dependent random fields, Statistical Inference Stochastic Processes, Vol. 5, 199-228.

[12] Doukhan, P., Louhichi, S. (1999) *A new weak dependence condition and applications to moment inequalities.* Stoch. Proc. Appl. 84, 313-342.

[13] Doukhan, P. and Wintenberger, O. (2007) A central limit theorem under non causal weak dependence and sharp moment assumptions, Prob. Math. Stat. 27, 45-73.

[14] Drost, F. C., Van Den Akker, R., Werker, B. J. (2007) *Note on integer valued bilinear time series models.* Statistics and Probability Letters.

[15] Du, J.-G., Li, Y. (1991) *The integer-valued autoregressive (INAR(p))* model. J. Times Ser. Anal. 12, 129-142.

[16] Efros, A. A., Leung, T. (1999) *Texture synthesis by non-parametric sampling.* In Proc. IEEE International Conference on Computer Vision 2 1033-1038. IEEE Computer Soc., Washington.

[17] Esary, J., Proschan, F., Walkup, D. (1967) *Association of random variables, with applications.* Ann. Math. Statist., 38, 5, pp. 1466-1474.

[18] Francq, C., Makarova, S., Zakoïan, J-M. (2008) *A class of stochastic unit-root bilinear processes. Mixing properties and unit-root test.* Forthcoming in the Journal of Econometrics.

[19] Francq, C., Zakoïan, J-M. *Inconsistency of the QMLE and asymptotic normality of the weighted LSE for a class of conditionally heteroscedastic models.* Working document. http ://www.eea-esem.com/files/papers/EEA-ESEM/2008/1099/statbprocesses 7juillet08.pdf

[20] Giraitis, L., Robinson, P., Surgailis, D. (2000) *A model for long memory conditional heteroscedasticity.* Annals of Applied Probability **10**-3, 1002-1024.

[21] Giraitis, L., Leipus, R., Robinson, P., Surgailis, D. (2004) *LARCH, Leverage, and Long Memory.* Journal of Financial Econometrics **2**-2, 177-210.

[22] Giraitis, L., Surgailis, D. (2002) *ARCH-type bilinear models with double long memory.* Stochastic Processes and their Applications **100**, 275-300.

[23] Guyon, X. (1995) *Random fields on a network : modeling, statistics, and applications.* Graduate Texts in Mathematics. Springer-Verlag, New-York.

[24] Helson, H., Lowdenslager, D. (1959) *Prediction theory and Fourier series in several variables.* Acta mathematica, 99, 165-202.

[25] Illig, A., Truong-Van, B. (2006)*Asymptotic results for spatial ARMA models.* Commun. Stat., Theory Methods 35, No.4, 671-688.

[26] Latour, A. (1998) *Existence and stochastic structure of a non-negative integervalued autoregressive process.* J. Time Ser. Anal. 19, 439Ũ455.

[27] Liu, W., Lin, Z. (2008) *Strong approximation for stationary processes.* Forthcoming in Stochastic Processes and their Applications.

[28] Loubaton, P. (1989) *Champs stationnaires au sens large sur* $\mathbb{Z}^2$ *: Propriétés structurelles et modèles paramétriques.* Traitement du signal, 6-4, 223-247.

[29] Paget, R., Longstaff, I. D. (1998) *Texture synthesis via a noncausal nonparametric multiscale Markov random field* IEEE Transactions on Image Processing, vol. 7, n 6, pp. 925-931.

[30] Rosenblatt, M. (1956) *A central limit theorem and a strong mixing condition.* Proc. Nat. Ac. Sc. U.S.A., 42, 43-47.

[31] Steutel, F., Van Harn, K. (1979) *Discrete analogues of selfdecomposability and stability.* Ann. Prob. 7, 893-899.

Chapitre 2

A fixed point approach to model random fields

Abstract

We introduce new models of stationary random fields, solutions of

$$X_t = F\left((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}}; \xi_t\right),$$

the input random field ξ is stationary, *e.g.* ξ is independent and identically distributed (iid). Such models extend most of those used in statistics. The (nontrivial) existence of such models is based on a contraction principle and Lipschitz conditions are needed ; those assumptions imply Doukhan and Louhichi (1999)'s [11] weak dependence conditions. In contrast to the concurrent ones, our models are not set in terms of conditional distributions. Various examples of such random fields are considered. We also use a very weak notion of causality of independent interest : it allows to relax the boundedness assumption of inputs for several new heteroscedastic models, solutions of a nonlinear equation.

Note

The content of this part is based on a paper, written in collaboration with Paul Doukhan.

2.1 Introduction

Description of random fields is a difficult task, a very deep reference is Georgii's (1988) book [11] ; a synthetic presentation is given by Föllmer (1985) [11]. The usual way to describe interactions makes use of conditional distributions with respect to large sets of indices. This presentation is natural for discrete valued random fields as in Comets *et alii* (2002), [1]. The existence of conditional densities is a more restrictive assumption for continuous state spaces. The existence of random fields

is often based on conditional specifications, see Föllmer (1985) ([11], pages 109-119) and Dobrushin (1970) [4], through Feller continuity assumptions. The uniqueness of Gibbs measures is often based on projective conditional arguments ; it follows with a mixing type argument. Such conditions rely on the regularity of conditional distributions ; applications to resampling exclude such hypotheses. Various applications to image, geography, agronomy, physic, astronomy or electromagnetism may for instance be considered, see [11] or [13].
We omit here any assumption relative to the conditional distributions. Our idea is to define random fields through more algebraic and analytic arguments. We present here the new models of stationary random fields subject to the relation :

$$X_t = F\left((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_t\right) \tag{2.1}$$

where $\xi = (\xi_t)_{t\in\mathbb{Z}^d}$ is an independent identically distributed (iid) random field. The independence of inputs ξ may also be relaxed to a stationarity assumption.
For the models with infinite interactions (6.4), the existence and uniqueness rely on the contraction principle. Lipschitz type conditions are thus needed, they are closely related to weak dependence, see [11]. Analogue weak dependence conditions are already proved in Shashkin (2005) for spin systems, [14]. A causal version of such models, random processes solutions of an equation $X_t = F(X_{t-1}, X_{t-2}, \ldots; \xi_t)$ $(t \in \mathbb{Z})$ is considered in Doukhan and Wintenberger (2006), [20] ; in this paper the results are proved in a completely different way fitting to coupling arguments. Our results state existence and uniqueness of a solution of (6.4) as a Bernoulli shift $X_t = H((\xi_{t-s})_{s\in\mathbb{Z}^d})$ as well as the weak dependence properties of this solution.
Our models are not necessarily Markov, neither linear or homoskedastic. Moreover the inputs do not need additional distributional assumptions (like for Gibbs random fields). They extend on ARMA random fields which are special linear random fields (see [13] or [16]). A forthcoming paper will be aimed at developing statistical issues of those models. Identification and estimation of random fields with integer values will be considered in [5].
The paper is organized as follows. We first recall weak dependence from [11] in § 2.2. General results are then stated for stationary (non necessarily independent) inputs. Those results imply heavy restrictions on the innovations in some cases : a convenient notion of causality is thus used. A last subsection addresses the problem of simulating such models.
A following section details examples of such models. They are natural extensions of the standard times series models. We shall especially consider LARCH(∞) and doubly stochastic linear random fields for which this causality allows to relax the boundedness assumptions. Proofs are postponed to a last section of the paper.

2.2 Main results

In order to state our dependence results, we first introduce the concepts of weak dependence. Our main results will be stated in the following subsection. After this, causality will be proved to imply other powerful results. A last subsection is aimed at describing a way to simulate those very general random fields.

2.2.1 Weak dependence

We recall here the weak dependence conditions introduced in Doukhan & Louhichi (1999). They may replace heavy mixing assumptions.

Definition 2.1 *Set* $\|(s_1,\ldots,s_d)\| = \max\{|s_1|,\ldots,|s_d|\}$ *for* $s_1,\ldots,s_d \in \mathbb{Z}$. *One* $E = \mathbb{R}^k$*−valued random field* $(X_t)_{t\in\mathbb{Z}^d}$ *is weakly dependent if for a sequence* $(\varepsilon(r))_{r\in\mathbb{N}}$ *with limit* 0

$$|Cov\,(f(X_{s_1},\ldots,X_{s_u})\,,\,g\,(X_{t_1},\ldots,X_{t_v})| \leqslant \psi(u,v,Lip\,f,Lip\,g)\varepsilon(r),$$

where indices $s_1,\ldots,s_u,\ t_1,\ldots,t_v \in \mathbb{Z}^d$ *are such that* $\|s_k - t_l\| \geq r$ *for* $1\leq k\leq u$ *and* $1\leq l\leq v$. *Moreover, the real valued functions* f, g *defined on* $\left(\mathbb{R}^k\right)^u$ *and* $\left(\mathbb{R}^k\right)^v$, *satisfy* $\|f\|_\infty$, $\|g\|_\infty \leqslant 1$ *and* $Lip\,f, Lip\,g<\infty$ *where a norm* $\|\cdot\|$ *is given on* $\mathbb{R}^k$ *and,*

$$Lip\,f = \sup_{(x_1,\ldots,x_u)\neq(y_1,\ldots,y_u)} \frac{|f(x_1,\ldots,x_u)-f(y_1,\ldots,y_u)|}{\|x_1-y_1\|+\cdots+\|x_u-y_u\|}.$$

If $\psi(u,v,a,b) = au+vb$, *this is denoted as* η*−dependence and the sequence* $\varepsilon(r)$ *will be written* $\eta(r)$.
If $\psi(u,v,a,b) = abuv$, *this is denoted as* κ*−dependence and the sequence* $\varepsilon(r)$ *will be written* $\kappa(r)$.
If $\psi(u,v,a,b) = au+vb+abuv$, *this is denoted as* λ*−dependence and the sequence* $\varepsilon(r)$ *will be written* $\lambda(r)$.

2.2.2 Random fields with infinite interactions

Let $\xi = (\xi_t)_{t\in\mathbb{Z}^d}$ be a stationary random field with values in E' (usually $E' = \mathbb{R}^{k'}$ for some $k'\geq 1$ but in some cases E' is a denumerable tensor product of such sets). We shall consider stationary $E = \mathbb{R}^k$ valued random fields driven by the implicit equation (6.4). For a topological space S, $\mathcal{B}(S)$ denote the Borel σ-algebra on S.
We denote $I = \mathbb{Z}^d\setminus\{0\}$. In the sequel, $F : \left(E^{(I)}\times E', \mathcal{B}(E^I)\otimes\mathcal{B}(E')\right) \to \left(E,\mathcal{B}(E)\right)$ denotes a measurable function defined for each sequence with a finite number of non-vanishing arguments ([1]). In this paper $\|\cdot\|$ will be arbitrary norms on E (or E' when needed). We will always use the suppremum

[1] If V denotes a vector space and B an arbitrary set then $V^{(B)}\subset V^B$ denotes the set of $v=(v_b)_{b\in B}$ such that there is some finite subset $B_1\subset B$ with $v_b = 0$ for each $b\notin B_1$.

norm on $\mathbb{Z}^d$ and this norm will be also denoted by $\|\cdot\|$. We prove that simple assumptions entail existence of a unique solution as a Bernoulli shift

$$X_t = H\left((\xi_{t-j})_{j\in\mathbb{Z}^d}\right)$$

Let μ denote ξ's distribution ; this is a probability measure on the measurable space $\left(E'^{\mathbb{Z}^d}, \mathcal{B}\big(E'^{\mathbb{Z}^d}\big)\right)$. For some $m \geqslant 1$, we denote $\|\cdot\|_m$ the usual norm of $\mathbb{L}^m$ and the space of μ-measurable H : $\big(E'^{\mathbb{Z}^d}, \mathcal{B}\big(E'^{\mathbb{Z}^d}\big)\big) \to (E, \mathcal{B}(E))$ with finite moments is denoted

$$\mathbb{L}^m(\mu) = \{H/\ \mathbb{E}\|H(\xi)\|^m < \infty\}.$$

We shall use the assumptions :

(H1) $\|F(0;\xi_0)\|_m < \infty$.

(H2) There exist constants $a_j \geqslant 0$, $j\in\mathbb{Z}^d$ such that $\forall z, z' \in E^{(\mathbb{Z}^d\setminus\{0\})}$,

$$\begin{aligned} \|F(z;\xi_0) - F(z';\xi_0)\| &\leq \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j\|z_j - z'_j\|, \quad a.s. \\ \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j &= e^{-\alpha} < 1. \end{aligned} \tag{2.2}$$

We now extend the function F to the trajectories of a stationary random field :

Lemma 2.1 *Assume* **(H1)** *and* **(H2)**. *Let X and X' be two $E-$valued stationnary random fields in $\mathbb{L}^m$, then :*

1) $\lim_{p\to\infty} F\big((X_j \mathbb{1}_{0<\|j\|\leq p})_{j\neq 0};\xi_0\big)$ *exists in $\mathbb{L}^m$ and a.s., we denote it* $F\left((X_j)_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_0\right)$.

2) $\left\|F\big((X_j)_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_0\big) - F\big((X'_j)_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_0\big)\right\|_m \leq \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j \left\|X_j - X'_j\right\|_m.$

Theorem 2.1 *Assume that ξ is stationary and* **(H1)** *and* **(H2)** *hold. Then there exists a unique stationary solution of equation (6.4). This solution writes $X_t = H\left((\xi_{t-j})_{j\in\mathbb{Z}^d}\right)$ for some $H\in\mathbb{L}^m(\mu)$.*

Lemma 2.8 below, will also provide us with an approximation of this solution with finitely many interactions.

Weak dependence of the solution (iid inputs)

In the general case we shall restrict to independent inputs to derive $\eta-$weak dependence of the previous solution.

Theorem 2.2 *Assume that ξ is iid and* **(H1)** *and* **(H2)** *hold. Then the stationary solution of equation (6.4) obtained in theorem 4.2 is $\eta-$weakly dependent and there exists a constant $C > 0$ with*

$$\eta(r) \leqslant C \cdot \inf_{p\in\mathbb{N}^*} \Big\{ e^{-\alpha \frac{r}{2p}} + \sum_{\|i\|>p} a_i \Big\}. \tag{2.3}$$

Remark. If $a_i = 0$ for $\|i\| > p$ then $\eta(r) \leqslant C \cdot e^{-\alpha \frac{r}{2p}}$.

Sub-geometric rates are now derived from specific decays of the coefficients :

Lemma 2.2 (Geometric decays) *If $a_i \leqslant Ce^{-\beta\|i\|}$ there exists a constant $C' > 0$ with*

$$\eta(r) \leq C' r^{\frac{d-1}{2}} e^{-\sqrt{\alpha\beta r/2}}.$$

Lemma 2.3 (Riemanian decays) *If $a_i \leq C\|i\|^{-\beta}$ for a $\beta > d$, there exists $C' > 0$ with*

$$\eta(r) \leq C' \left(\frac{r}{\ln r}\right)^{d-\beta}.$$

Thus a large range of decay rates may be considered for such models of random fields.

Weak dependence of the solution (dependent inputs)

If ξ is either η or $\lambda-$dependent it may be proved in specific examples that weak dependence is hereditary. Here follows a general result. The following assumption will be necessary :

(H2') There exist a subset $\Xi \subset E'$ with $\mathbb{P}(\xi_0 \in \Xi) = 1$, nonnegative constants with $\sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j = e^{-\alpha} < 1$ and a constant $b > 0$ such that

$$\|F(x;u) - F(x';u')\| \leq \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j \|x_j - x'_j\| + b\left\|u - u'\right\|,$$

for all $x, x' \in E^{(\mathbb{Z}^d\setminus\{0\})}$ and $u, u' \in \Xi$.

We quote that assumption **(H2')** is more restrictive than **(H2)**

Proposition 2.1 *Assume* **(H1)** *and* **(H2')**.

1) *If the random field ξ is $\eta-$weakly dependent, with weak dependence coefficients $\eta_\xi(r)$, then X is $\eta-$weakly dependent with, for some $C > 0$,*

$$\eta(r) \leq C \inf_{p\in\mathbb{N}^*} \Big\{ \sum_{\|j\|>p} a_j + \inf_{n\in\mathbb{N}^*} \Big\{ a^n + p^n \eta_\xi\left((r - 2pn) \vee 0\right) \Big\}\Big\}.$$

2) *If the random field ξ is $\lambda-$weakly dependent, with dependence coefficients denoted $\lambda_\xi(r)$, then X is $\lambda-$weakly dependent with, for some $C>0$,*

$$\lambda(r) \le C \inf_{p\in\mathbb{N}^*} \left\{ \sum_{\|j\|>p} a_j + \inf_{n\in\mathbb{N}^*} \left\{ a^n + p^{2n}\lambda_\xi\left((r-2pn)\vee 0\right) \right\} \right\}.$$

Remark. For models with finite interactions, *i.e.* $F(x;u)=f(x_{j_1},\dots,x_{j_k};u)$ for $x=(x_j)_{j\neq 0}$, this simply writes

$$\begin{aligned} \eta(r) &\le c \inf_{n\in\mathbb{N}^*} \left\{ a^n + k^n \eta_\xi\left((r-2\rho n)\vee 0\right) \right\}, \\ \lambda(r) &\le c \inf_{n\in\mathbb{N}^*} \left\{ a^n + k^{2n} \lambda_\xi\left((r-2\rho n)\vee 0\right) \right\}, \end{aligned}$$

here $\rho=\max\{\|j_1\|,\dots,\|j_k\|\}$. If $\eta_\xi(r)$ or $\lambda_\xi(r)$ have geometric or Riemannian decay the same holds for the output random field. More precisely set $a=e^{-\alpha}$ and $k=e^\kappa$ under η-dependence and $k^2=e^\kappa$ under λ-dependence, then decay rates of the outputs (X_t) write

Geometric decays :	$e^{-\frac{\alpha\beta}{\alpha+2\rho\beta+\kappa}r}$,	for dependence decays of the inputs with order	$e^{-\beta r}$,
Riemannian decays :	$r^{-\frac{\alpha b}{\alpha+\kappa}}$,	for dependence decays of the inputs with order	r^{-b}.

2.2.3 Causality

For $d=1$, the recurrence equation $X_t=\xi_t(a+bX_{t-1})$ is given with $F(x;u)=u(a+bx_1)$. There exist a stationary solution with ξ_t and X_{t-1} independent. Here **(H2)** implies that innovations are bounded, which seems unrealistic. In this example, instead of $H\left((\xi_t)_{t\in\mathbb{Z}}\right)\in\mathbb{L}^m(\mu)$, this is enough to exhibit solutions $H\left((\xi_t)_{t\geqslant 0}\right)\in\mathbb{L}^m(\mu)$ (which is independent of $(\xi_s)_{s<0}$). This allows to replace *suprema* by *integrals* in **(H2)** in order to derive a contraction principle. Causality of random fields has been considered in Helson and Lowdenslager (1959) [12] ; we adapt this idea in order to relax the previous assumption.

Definition 2.2 (causality) *If $A\subset\mathbb{Z}^d\setminus\{0\}$, we denote $c(A)$ the convex cone of $\mathbb{R}^d$ generated by A,*

$$c(A)=\left\{ \sum_{i=1}^k r_i j_i \Big/ \ (j_1,\dots,j_k)\in A^k, (r_1,\dots,r_k)\in\mathbb{R}_+^k,\ k\ge 1 \right\}.$$

1) The set A is a causal subset of $\mathbb{Z}^d$ if $c(A)\cap\big(-c(A)\big)=\{0\}$.
2) If F is measurable with respect to the σ-algebra $\mathfrak{F}_A\otimes\mathcal{B}(E')$ for some causal set A, then the equation $X_t=F\left((X_{t-j})_{j\in I};\xi_t\right)$ is A-causal.

For a causal set $A\subset\mathbb{Z}^d$, we denote by $\widetilde{A}$ the subset $c(A)\cap\mathbb{Z}^d$.

Examples. A singleton is causal, as well as $\{i,j\}$ if and only if $-j \notin i \cdot \mathbb{R}^+$. The half plane $\{(i,j) \in \mathbb{Z}^2 / i > 0\} \bigcup \{(0,j); j > 0\} \subset \mathbb{Z}^2$ is also causal.

One consequence of this notion is the elementary lemma :

Lemma 2.4 *If A is a causal subset of $\mathbb{Z}^d$, then $\forall (j,j') \in A \times \widetilde{A}$ we have $j + j' \neq 0$.*

For a linear basis $b = (b_1, \ldots, b_d)$ of $\mathbb{R}^d$, $(x_1, \ldots, x_d) \mapsto x_1 b_1 + \cdots + x_d b_d$, defines an isomorphism $f : \mathbb{R}^d \to \mathbb{R}^d$. We denote by $\leq_b$ the total order relation on $\mathbb{R}^d$ defined by :

$$u \leq_b v \Leftrightarrow f^{-1}(u) \leq_{lex} f^{-1}(v)$$

with $\leq_{lex}$ the lexicographic order on $\mathbb{R}^d$.

Proposition 2.2 (characterization of causal sets) *If B is a convex cone of $\mathbb{R}^d$ such that $B \cap (-B) = \{0\}$ there exists a basis b of $\mathbb{R}^d$ such that $B \subset \{j \in \mathbb{R}^d / 0 \leq_b j\}$. Moreover if b is a basis of $\mathbb{R}^d$, $\{j \in \mathbb{Z}^d / 0 <_b j\}$ is a causal set of $\mathbb{Z}^d$ witch will be called maximal causal subset.*

Remarks.

- The maximal causal subsets of $\mathbb{Z}$ are $\{1, 2, 3, \ldots\}$ and $\{-1, -2, \ldots\}$. An example of maximal causal subset of $\mathbb{Z}^2$ is $\{(i,j) \in \mathbb{Z}^2 / i > 0 \text{ or } (i = 0, j > 0)\}$.
- Helson and Lowdenslager (1959) [12] define symmetric half planes as subsets $S \subset \mathbb{Z}^2$ such that S is stable by addition and $S \cup (-S) = \mathbb{Z}^2$, $S \cap (-S) = \{0\}$. A nice review of this causality condition is given in Loubaton (1989) [13], applications are essentially given in terms of linear random fields. Note that $S \backslash \{0\}$ is a maximal causal subset of $\mathbb{Z}^2$. This notion plays a prominent part in prediction theory of 2-D stationary process (see [13]).

If $D \subset \mathbb{Z}^d$, we denote by π_s (respectively π'_s) the coordinate applications in E^{Z^d} (resp. in $(E')^{\mathbb{Z}^d}$), $\mathfrak{F}_D = \sigma(\pi_s; s \in D)$ and $\mathfrak{F}'_D = \sigma(\pi'_s; s \in D)$. Hence we denote by $\mathbb{L}^m_D(\mu)$ the subspace of $\mathbb{L}^m(\mu)$ of functions μ-measurable with respect to $\mathfrak{F}'_D$. The following result takes this definition into account to relax the assumptions in theorem 4.2,

Theorem 2.3 *Let $X_t = F\left((X_{t-j})_{j \in \mathbb{Z}^d \backslash \{0\}}; \xi_t\right)$ be a A-causal equation with iid inputs ξ. Besides the assumption* **(H1)** *we assume the following condition :*
(H3) *there exist nonnegative constants with $\sum_{j \in A} a_j = e^{-\alpha} < 1$ and*

$$\|F(x;\xi_0) - F(x';\xi_0)\|_m \leq \sum_{j \in A} a_j \|x_j - x'_j\|, \qquad \forall x, x' \in E^{(\mathbb{Z}^d \backslash \{0\})}.$$

Then there exists a unique strictly stationary solution X of this equation in $\mathbb{L}^m$ if for each $t \in \mathbb{Z}^d$, X_t is measurable wrt $\sigma\left(\xi_{t-j} / j \in \widetilde{A}\right)$.
This solution writes $X_t = H\left((\xi_{t-j})_{j \in \mathbb{Z}^d}\right)$ where $H \in \mathbb{L}^m_{\widetilde{A}}$. and it is $\eta-$weakly dependent; moreover relation (2.3) still holds for a constant $C > 0$.

Now the function F is extended as follows :

Lemma 2.5 *Suppose* **(H1)** *and* **(H3)**. *If* ξ_0 *is independent of* $\sigma\big((X_j, X'_j)/j \in A\big)$ *for two random fields* X *and* X' *in* $\mathbb{L}^m$ *then,*

1) $\lim_{p\to\infty} F\big((X_j \mathbb{1}_{0<\|j\|\le p})_{j\neq 0};\xi_0\big)$ *exists in* $\mathbb{L}^m$ *and it is denoted* $F\big((X_j)_{j\neq 0};\xi_0\big)$.
2) $\left\|F\big((X_j)_{j\neq 0};\xi_0\big) - F\big((X'_j)_{j\neq 0};\xi_0\big)\right\|_m \le \sum_{j\in A} a_j \left\|X_j - X'_j\right\|_m$.

2.2.4 Simulation of the model

Simulations of those models are deduced from the proof of the existence theorems based on the Picard fixed point theorem. Consider the shift operators $\theta_j : \big(E'\big)^{\mathbb{Z}^d} \to \big(E'\big)^{\mathbb{Z}^d}$ defined as $(x_k)_{k\in\mathbb{Z}^d} \mapsto (x_{k+j})_{k\in\mathbb{Z}^d}$. For $H \in \mathbb{L}^m(\mu)$ we note

$$\Phi_p(H) = F\left(\left((H\circ\theta_j)\mathbb{1}_{\|j\|\le p}\right)_j ; \pi_0\right)$$

It is shown in theorem 4.2's proof that the application $\Phi : \mathbb{L}^m(\mu) \to \mathbb{L}^m(\mu)$ given by

$$\Phi(H) = F((H\circ\theta_j)_{j\neq 0};\pi_0).$$

is well defined and has a fixed point in $\mathbb{L}^m(\mu)$.
The proof of theorem 2.3 shows that it is also the case for a $A-$causal equation if we replace $\mathbb{L}^m(\mu)$ by $\mathbb{L}^m_{\widetilde{A}}(\mu)$.
For $n, p \in \mathbb{N}^*$, $t \in \mathbb{Z}^d$ we denote $X^n_t = \Phi^{(n)}(0)\left((\xi_{t-j})_{j\in\mathbb{Z}^d}\right)$ and $X^n_{p,t} = \Phi^{(n)}_p(0)\left((\xi_{t-j})_{j\in\mathbb{Z}^d}\right)$.

Lemma 2.6 *We assume that conditions in theorem 4.2 or in theorem 2.3 hold for some* $m \geqslant 1$. *Let* $n \in \mathbb{N}$ *then :*

1. *For every* $t \in \mathbb{Z}^d$, $\|X_t - X^n_t\|_m \le a^n \|X_0\|_m$, *hence* $\lim_{n\to\infty} X^n_t = X_t$ *a.s.*

2. *if* $p \in \mathbb{N}$ *we have,* $\left\|X_t - X^n_{p,t}\right\|_m \le \|X_0\|_m \left\{a^n + \frac{1}{1-a}\sum_{\|j\|>p} a_j\right\}$. *Thus if* $p = p_n$ *is chosen such that* $\sum_{n\geqslant 1}\Big(\sum_{\|j\|>p_n} a_j\Big)^m < \infty$ *then*

$$\lim_{n\to\infty} X^n_{p_n,t} = X_t, \quad a.s. \tag{2.4}$$

Remarks.
– If the random field has finitely many interactions, then 1. provides a simulation scheme.
– For each finite p the operator Φ_p can be calculated thus relation (2.4) provides an explicit simulation scheme even for infinitely many interactions.

- *A.s.* convergence rates may also be evaluated in the previous lemma. They write $o_{a.s.}(a^n n^\varepsilon)$ in the first point for each $\varepsilon > 1/m$ and $o_{a.s.}(n^{-\varepsilon})$ for $0 < \varepsilon < \alpha - 1/m$ if $\sum_{\|j\|>p_n} a_j \leqslant Cn^{-\alpha}$ for some $C > 0, \alpha > 1/m$ in the point 2.
- If $T \subset \mathbb{Z}^d$ is a finite set the random field X may be analogously simulated over T and $(X_t)_{t\in T}$ is estimated by $\left(X^n_{p_n,t}\right)_{t\in T}$.

Simulation scheme for finitely many interactions

Let $F(x;u) = f(x_{j_1}, \ldots, x_{j_k}; u)$. The sequence of random fields X^n is defined from :

$$X_t^1 = f(0;\xi_t), t \in \mathbb{Z}^d, \quad X_t^{n+1} = f\big(X^n_{t-j_1}, \ldots, X^n_{t-j_k}; \xi_t\big), \qquad \text{for } n \geqslant 0$$

We now simulate samples $(X_t^{10})_{1\le t_1,t_2\le 15}$ of LARCH models with $d = 2$, $k = k' = 1$ and $p = 10$:

$$X_t = \xi_t\Big(1 + \sum_{0<\|j\|\le p} a_j X_{t-j}\Big)$$

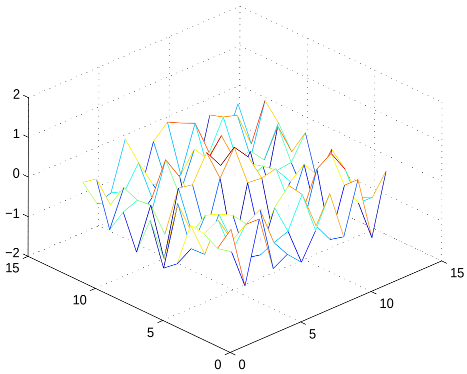


FIG. 2.1: Non causal LARCH field

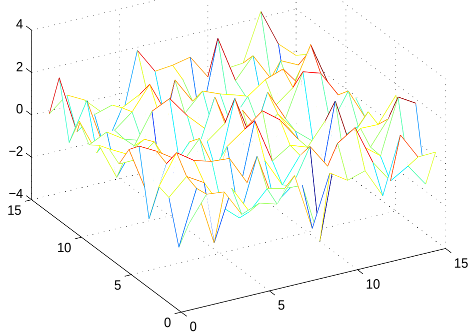


FIG. 2.2: Causal LARCH field

1) In the figure 7.2, we represent the non causal case with $a_j = \dfrac{0.05}{j_1^2 + j_2^2}$ and ξ_0 is uniform on $[-1, 1]$.

2) Figure 5.2 deals with the causal case with $a_j = \dfrac{0.05}{j_1^2 + j_2^2}$ if $0 \le j_1, j_2 \le 10$ and $a_j = 0$ otherwise. In this case, ξ_0 is $N(0,1)$-distributed.

2.3 Examples

Theorems 2.2 and 2.3 are now applied to examples of random fields with infinite interactions. Causality will allow to weaken moment conditions. In fact, theorem 2.2 proves a contraction principle in $\mathbb{L}^m$ for each value of m while theorem 2.3 only works with one fixed value of m.

2.3.1 Finite interactions random fields

If $\xi_t = (\zeta_t, \gamma_t)$ with $\zeta_t \in \mathbb{R}^p$ and γ_t a $p \times q$ matrix, and functions $f(\cdot) \in \mathbb{R}^p$ and $g(\cdot) \in \mathbb{R}^q$

$$X_t = f(X_{t-\ell_1}, \ldots, X_{t-\ell_k}) + \gamma_t g(X_{t-\ell_1}, \ldots, X_{t-\ell_k}) + \zeta_t \tag{2.5}$$

with $\ell_1, \ldots, \ell_k \neq 0$.
E.g. non linear auto-regression corresponds to $\gamma_t \equiv 0$ and ARCH type models are obtained with $\zeta_t = 0$ (classically $p = q = 1$, f is linear and $g^2(x_1, \ldots, x_k)$ is an affine function of $x_1^2, \ldots, x_k^2$).
Theorems 4.2, 2.2, 2.3 imply the following lemma,

Corollary 2.1 *Suppose* $\|\zeta_0\|_m < \infty$ *and*

$$\left\{ \begin{array}{lll} \|f(x_1, \ldots, x_k) - f(y_1, \ldots, y_k)\| & \leq & \sum_{i=1}^k b_i \|x_i - y_i\|, \\ \|g(x_1, \ldots, x_k) - g(y_1, \ldots, y_k)\| & \leq & \sum_{j=1}^k c_j \|x_j - y_j\| \end{array} \right. .$$

1. *If ξ is iid and* $\displaystyle\sum_{i=1}^k \left(b_i + \|\gamma_0\|_\infty c_i\right) = e^{-\alpha} < 1$*, then* $\eta(r) \leqslant C \left(e^{-\frac{\alpha}{2k}}\right)^r$ *for model (2.5).*

   If the equation (2.5) is causal and $\displaystyle\sum_{i=1}^k \left(b_i + \|\gamma_0\|_m c_i\right) = e^{-\alpha} < 1$*, the same holds.*

2. *If now ξ is η or λ-weakly dependent, g bounded and* $\displaystyle\sum_{i=1}^k \left(b_i + \|\gamma_0\|_\infty c_i\right) = e^{-\alpha} < 1$*, then X is η or λ-weakly dependent. Decays are given according to proposition 2.1.*

The remark following proposition 2.1 states precise decays. The volatility coefficients γ_t need to be bounded in the general case and they only have finite moments under causality. Functions f and g may only depend on a strict subset of the indices $1, \ldots, k$.

2.3.2 Linear fields

Let X be a solution of the equation

$$X_t = \sum_{j \in A} \alpha_t^j X_{t-j} + \zeta_t, \tag{2.6}$$

innovations ζ_t are vectors of $E=\mathbb{R}^k$ and coefficients α_t^j are $k\times k$ matrices, $\|\cdot\|$ is a norm of algebra on this set of matrices and X will be an E valued random field. Let $A\subset\mathbb{Z}^d\setminus\{0\}$, we assume that the iid random field $\xi=\Big((\alpha_t^j)_{j\in A},\zeta_t\Big)_{t\in\mathbb{Z}^d}$ takes now its values in $(M_{k\times k})^A\times E$; here $M_{k\times k}$ denotes the set of $k\times k$ matrices.

Proposition 2.3 *If $b=\sum_{j\in A}\|\alpha_0^j\|_\infty<1$, then theorem 2.2 applies with $a_j=\|\alpha_0^j\|_\infty$.*
For a causal equation if $b=\sum_{j\in A}\|\alpha_0^j\|_m<1$ theorem 2.3 applies with $a_j=\|\alpha_0^j\|_m$.
In both cases the solution of equation (6.8) writes a.s. and in $\mathbb{L}^m$,

$$X_t=\zeta_t+\sum_{j\in A}\alpha_t^j\xi_{t-j}+\sum_{i=2}^\infty\sum_{j_1,\dots,j_i\in A}\alpha_t^{j_1}\alpha_{t-j_1}^{j_2}\cdots\alpha_{t-j_1-\cdots-j_{i-1}}^{j_i}\zeta_{t-(j_1+\cdots+j_i)}.$$

This means that the random coefficients are bounded in the general case and they need only to have have finite moments under causality.

Examples. If the sequence $(\alpha_t^j)_t$ is deterministic then those models extend on linear auto-regressive models. If only a finite number of coefficients α_t^j do not vanish we obtain auto-regressive models with random coefficients, see [15].

2.3.3 LARCH(∞) random fields

Stationary innovations ξ_t are now $k\times k'$ matrices and $\|\cdot\|$ will denote a norm $k\times k'$ or $k'\times k$ matrices while $X_t\in\mathbb{R}^k$. For bounded innovations we first recall

Theorem 2.4 (Doukhan, Teyssière, Winant (2006)) *Let α_j be a $k'\times k$ matrix for $j\in\mathbb{Z}^d\setminus\{0\}$, note $A(x)=\sum_{\|j\|\geq x}\|\alpha_j\|$ and suppose that $\lambda=A(1)\|\xi_0\|_\infty<1$, then*

$$X_t=\xi_t\left(a+\sum_{k=1}^\infty\sum_{j_1,\dots,j_k\neq0}\alpha_{j_1}\xi_{t-j_1}\cdots\alpha_{j_k}\xi_{t-j_1-\cdots-j_k}a\right)\tag{2.7}$$

is a solution of the equation

$$X_t=\xi_t\left(a+\sum_{j\neq0}\alpha_jX_{t-j}\right),\quad t\in\mathbb{Z}^d\tag{2.8}$$

if moreover ξ is iid, then

$$\eta(r)\leqslant\mathbb{E}\|\xi_0\|\left(\mathbb{E}\|\xi_0\|\sum_{k<r/2}\lambda^{k-1}A\left(\frac{r}{k}\right)+\frac{\lambda^{[r/2]}}{1-\lambda}\right)\|a\|.$$

If we use theorem 2.2 we also obtain that eqn. (6.10) admits a unique Bernoulli shift $\mathbb{L}^m$ solution. Note that this solution is bounded. Notice that for Riemannian decay the previous $A(u) \leqslant Cu^{-c}$ relation yields $\eta(r) = \mathcal{O}(r^{-c})$ while theorem 2.3 only provides us this bound up to a log-loss ; geometric decays yield the same result for both cases.
Bounded innovations ξ_t look unnatural hence we investigate below the causal case. Let A a causal subset of $\mathbb{Z}^d$ and

$$X_t = \xi_t \left(a + \sum_{s \in A} a_s X_{t-s} \right) \tag{2.9}$$

Proposition 2.4 *If $b\|\xi_0\|_m < 1$ with $b = \sum_{s\in A} \|a_s\|$, theorem 2.3 applies with $a_j = \|\xi_0\|_m \|\alpha_j\|$ to the solution (2.7) of eqn. (2.9) (we set $\alpha_j = 0$ for $j \notin A$).*

2.3.4 Non linear ARCH(∞) random fields

Models with

$$X_t = \xi_t \Big(a + \sum_{j \neq 0} g_j(X_{t-j}) \Big)$$

clearly extend on LARCH(∞) models ; bounded functions g_j provide robust models.

Corollary 2.2 *If $\|g_j(x) - g_j(y)\| \leq \alpha_j \|x - y\|$ and $\|\xi_0\|_\infty \sum_{j\neq 0} \alpha_j < 1$, theorem 2.2 holds with $a_i = \|\xi_0\|_\infty \alpha_i$ (innovations are bounded here).*
Assume now that $g_i \equiv 0$ for $i \notin A$, causal set then theorem 2.3 holds with $a_i = \|\xi_0\|_m \alpha_i$ (and now the innovations do not need anymore to be bounded).

This causality argument improves on [7] by only assuming finite moments for innovations instead of boundedness.

2.3.5 Mean field type model

Consider innovations in $\mathbb{R}^{k'}$ and $k \times k$ matrices α_i,

$$X_t = f\Big(\xi_t, \sum_{s \neq t} \alpha_{s-t} X_s \Big) \tag{2.10}$$

Corollary 2.3 *Assume that $f : \mathbb{R}^{k'} \times \mathbb{R}^k \to \mathbb{R}^k$ satisfies*

$$\sup_{u \in \mathbb{R}^{k'}} \|f(u, x) - f(u, y)\| \leq b\|x - y\|, \ \forall x, y \in \mathbb{R}^k, \qquad b \sum_{i \neq 0} \|\alpha_i\| < 1.$$

then equation (2.10) admits a unique solution in $\mathbb{L}^m$ written as a Bernoulli shift and this solution is $\eta-$weakly dependent with $a_i = b\|\alpha_i\|_1$.
The same results hold if now $a_i = 0$ for $i \notin A$ with A is causal in $\mathbb{Z}^d$ and,

$$\big\|f\big(\xi_0, x)\big) - f\big(\xi_0, y\big)\big\|_m \leq b\|x-y\|, \ \forall x, y \in \mathbb{R}^k, \qquad b\sum_{i\neq 0}\|\alpha_i\| < 1.$$

LARCH(∞) models take this form.

2.4 Proofs

We begin with the proof of some lemmas which relate the assumptions to contraction conditions in the space of Bernoulli shifts. Then we give separated proofs for existence and weak dependence properties. Those proofs always follow two steps since we first consider models with a finite range. For shortness we write here $I = \mathbb{Z}^d \setminus \{0\}$.

2.4.1 Proof of lemma 2.1

For $p \in \mathbb{N}^*$, we set $Y_p = F\big((X_j 1\!\!1_{0<\|t\|\leq p})_j, \xi_0\big)$ and $Y'_p = F\big((X'_j 1\!\!1_{0<\|t\|\leq p})_j, \xi_0\big)$.

1. If $q \in \mathbb{N}^*$ from assumption (H2),

$$\|Y_p - Y_{p+q}\| \leq \sum_{p<\|j\|\leq p+q} a_j \|X_j\|, \qquad a.s.$$

   Since the serie $\sum_{j\in I} a_j \|X_j\|_m$ is convergent the serie $\sum_{j\in I} a_j \|X_j\|$ converges *a.s.* Hence, we deduce that a.s $(Y_p)_{p\in\mathbb{N}^*}$ is a Cauchy sequence in E and then converges. We denote by $Y = F\big((X_j)_{j\neq 0}; \xi_0\big)$ this limit.
   Moreover, for $p \in \mathbb{N}^*$, we have :

$$\|Y_p - F(0;\xi_0)\|_m \leq \sum_{0<\|j\|\leq p} a_j \|X_j\|_m$$

   This proves that $Y_p \in \mathbb{L}^m$. Hence the convergence in $\mathbb{L}^m$ is a simply consequence of the Fatou lemma since :

$$\|Y - Y_p\|_m \leq \liminf_{q\to\infty} \|Y_q - Y_p\|_m \leq \sum_{\|j\|>p} a_j \|X_0\|_m$$

2. If $p \in \mathbb{N}^*$, we have using **(H2)** : $\left\|Y_p - Y'_p\right\|_m \leq \sum_{j\neq 0} a_j \left\|X_j - X'_j\right\|_m$,
   hence the result follows with $p \to \infty$. □

2.4.2 Proof of the existence theorem 4.2

Assuming **(H1)** and **(H2)** we set $a = \sum_{j\neq 0} a_j$. With the notations of paragraph 2.4.,

$$\Phi_p(H) = F\left(\left((H\circ\theta_j)\mathbb{1}_{\|j\|\le p}\right)_j; \pi_0\right), \ \forall H \in \mathbb{L}^m(\mu)$$

As a direct consequence of lemma 2.1, $\lim_{p\to\infty}\Phi_p(H)$ exists in $\mathbb{L}^m(\mu)$. Denote this limit by $F((H\circ\theta_j)_{j\neq 0};\pi_0)$, the application $\Phi : \mathbb{L}^m(\mu)\to\mathbb{L}^m(\mu)$ is defined as

$$\Phi(H) = F\left((H\circ\theta_j)_{j\neq 0};\pi_0\right).$$

Let show that Φ is a contraction of $\mathbb{L}^m(\mu)$.If H, $H'\in\mathbb{L}^m(\mu)$, then applying the lemma 2.1 to the random fields X and X' defined as $X_j = H\circ\theta_j(\xi)$ and $X'_j = H'\circ\theta_j(\xi)$, we obtain :

$$\begin{aligned}\|\Phi(H)(\xi)-\Phi(H')(\xi)\|_m &\le \sum_{j\neq 0} a_j\|H\circ\theta_j(\xi)-H'\circ\theta_j(\xi)\|_m\\ &\le \sum_{j\neq 0} a_j\|H(\xi)-H'(\xi)\|_m\end{aligned}$$

Picard fixed point theorem applies since the space $\mathbb{L}^m(\mu)$ is complete. There exists a unique $H\in\mathbb{L}^m(\mu)$ with $\Phi(H)=H$ thus $H(\xi) = F\left((H\circ\theta_j(\xi))_{j\in\mathbb{Z}^d};\xi_0\right)$, *a.s.* Set $X_t = H\left((\xi_{t-i})_{i\in\mathbb{Z}^d}\right)$ then with stationarity of ξ and since $\mathbb{Z}^d$ is denumerable we get

$$X_t = F\left((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_t\right), \quad \forall t\in\mathbb{Z}^d \qquad a.s.$$

Let Y be a stationary solution of this equation, we denote $u_t = \|X_t - Y_t\|_1$ for each $t\in\mathbb{Z}^d$. We obtain $u_t\le\sum_{j\neq 0} a_j u_{t-j}$. As $\sup_t u_t \le \|X_0\|_1+\|Y_0\|_1<\infty$ we note that the previous relation implies $\sup_t u_t \le a\sup_t u_t$. Hence $u_t = 0$ for each t. Thus $X_t = Y_t$ *a.s* for each t. $\square$

2.4.3 Proof of theorem 2.2

For an independent copy $(\xi'_t)_{t\in\mathbb{Z}^d}$ of $\xi=(\xi_t)_{t\in\mathbb{Z}^d}$ and $s\in\mathbb{R}^+$, we set $\xi^{(s)} = (\xi_t^{(s)})_{t\in\mathbb{Z}^d}$ with $\xi_t^{(s)} = \xi_t$ if $\|t\|<s$ and $\xi_s^{(s)} = \xi'_s$ else. For a Bernoulli shift defined by H a straightforward extension of a result in [11] to random fields implies

$$\eta(r)\le 2\delta_{r/2}, \qquad \text{where} \qquad \delta_r = \left\|H(\xi) - H\big(\xi^{(r)}\big)\right\|_1 \tag{2.11}$$

Weak dependence under finite interactions

We first assume that F depends finitely many variables

$$X_t = F(X_{t-j_1},\ldots,X_{t-j_k};\xi_0)$$

Lipschitz coefficients of F in condition **(H2)** write $a_1, \dots, a_k$ and we set $a = \sum_{i=1}^k a_i < 1$ and $\rho = \max\{\|j_1\|, \dots, \|j_k\|\}$. Let H be the element of $\mathbb{L}^m(\mu)$ with $X_t = H\left((\xi_{t-i})_{i\in\mathbb{Z}^d}\right)$ and $\delta_r = \mathbb{E}\left\|H(\xi) - H(\xi^{(r)})\right\| = \mathbb{E}\left\|X_0 - X_0^{(r)}\right\|$ with $X_t^{(r)} = H((\xi_{t-i}^{(r)})_{i\in\ \mathbb{Z}^d})$.

Lemma 2.7 *Assume that* **(H1)** *and* **(H2)** *hold, then* $\delta_r \le 2\|X_0\|_1 a^{\frac{r}{\rho}}$ *hence* $\delta_r \to_{r\to\infty} 0$.

Proof of lemma 2.7. Let $r > 0$. Since ξ and $\xi^{(r)}$ admit the same distribution, we have for each t :

$$X_t = F(X_{t-j_1}, \dots, X_{t-j_k}; \xi_t),\ X_t^{(r)} = F(X_{t-j_1}^{(r)}, \dots, X_{t-j_k}^{(r)}; \xi_t^{(r)})$$

If $\|t\| < r$ then $\xi_t^{(r)} = \xi_t$ and using **(H2)**, we have :

$$\|X_t - X_t^{(r)}\|_1 \le \sum_{l=1}^k a_l \left\|X_{t-j_l} - X_{t-j_l}^{(r)}\right\|_1 \tag{2.12}$$

Set now $i = -[-\frac{r}{\rho}]$ if $r \ge \rho$, then if $u \le i-1$ and $l_1, \dots, l_u \in \{1, \dots, k\}$: $\|j_{l_1} + j_{l_2} + \cdots + j_{l_u}\| < r$. We use inequality (2.12) to derive recursively the bounds

$$\begin{aligned}\left\|X_0 - X_0^{(r)}\right\|_1 &\le \sum_{l_1=1}^k a_{l_1} \sum_{l_2=1}^k a_{l_2} \cdots \sum_{l_i=1}^k a_{l_i} \left\|X_{-(j_{l_1}+j_{l_2}+\cdots+j_{l_i})} - X^{(r)}_{-(j_{l_1}+j_{l_2}+\cdots+j_{l_i})}\right\|_1 \\ &\le 2\|X_0\|_1 a^i\end{aligned}$$

From $i \ge r/\rho$ we get $\|X_0 - X_0^{(r)}\|_1 \le 2\|X_0\|_1\, a^{\frac{r}{\rho}}$ thus $\delta_r \le 2\|X_0\|_1 a^{\frac{r}{\rho}}$.
If now $r < \rho$, $\left\|X_0 - X_0^{(r)}\right\|_1 \le \sum_{l_1=1}^k a_{l_1} \left\|X_{-j_{l_1}} - X_{-j_{l_1}}^{(r)}\right\|_1$.
Thus $\delta_r \le 2\|X_0\|_1 a^i \le 2\|X_0\|_1\, a^{\frac{r}{\rho}}$. The result follows with $a < 1$. □

We now set a useful result. $(X_t)_{t\in\mathbb{Z}^d}$ and $(X_{p,t})_{t\in\mathbb{Z}^d}$ will denote for $p \geqslant 0$ the previous unique solution of the equations (6.4) and $Z_t = F\left((Z_{t-j}1_{\{0<\|j\|\le p\}})_{j\in\mathbb{Z}^d\setminus\{0\}}; \xi_t\right)$.

Lemma 2.8 *Assume that the conditions in theorem 4.2 hold. Then* $X_{p,t} \to_{s\to\infty} X_t$ *in* $\mathbb{L}^m$, *for each* $t \in \mathbb{Z}^d$.

Proof. $\|X_{p,0} - X_0\|_m \le a\|X_{p,0} - X_0\|_m + \|X_0\|_m \sum_{\|j\|>p} a_j$, thus $\|X_{p,0} - X_0\|_m \le \frac{\sum_{\|j\|>p} a_j}{1-a}\|X_0\|_m$ which entails the first result. We also quote that $\sup_p \|X_{p,0}\|_m < \infty$. □

Weak dependence

Lemma 2.9 *Assume that the conditions in theorem 4.2 hold. Then the random field* $(X_t)_{t\in\mathbb{Z}^d}$ *is* η*-weakly dependent.*

Proof. Recall that $\sup_p \|X_{p,0}\|_m < \infty$; if $m \geqslant 1$, weak dependence follows from

$$\begin{aligned}\mathbb{E}\|X_0^{(r)} - X_0\| &\leq \mathbb{E}\|X_0^{(r)} - X_{p,0}^{(r)}\| + \mathbb{E}\|X_{p,0}^{(r)} - X_{p,0}\| + \mathbb{E}\|X_{p,0} - X_0\| \\ &= 2\mathbb{E}\|X_{p,0} - X_0\| + \mathbb{E}\|X_{p,0}^{(r)} - X_{p,0}\|\end{aligned}$$

For $r \geq p$, from lemma 2.7 we derive $\mathbb{E}\|X_{p,0}^{(r)} - X_{p,0}\| \leq 2\|X_{p,0}\|_1 \Big(\sum_{\|j\| \leq p} a_j \Big)^{\frac{r}{p}}$. Hence $\delta_r = \mathbb{E}\|X_0^{(r)} - X_0\| \leq 2 \cdot \frac{\|X_0\|_1}{1-a} \sum_{\|j\|>p} a_j + 2\|X_{p,0}\|_1 \, a^{\frac{r}{p}}$.
With $\sup_p \|X_{p,0}\|_1 < \infty$ there exists $C > 0$ with $\delta_r \leq C \cdot \inf_p \Big\{ \sum_{\|j\|>p} a_j + a^{\frac{r}{p}} \Big\}$. Using (2.11) we prove that $(X_t)_t$ is η-weakly dependent and $\eta(r) \leq \delta_{r/2}$. □

Decay rates

Using the representation of the solution as a Bernoulli shift and the inequality (2.11) this will be enough to bound the expression of δ_r. Set $b_p = \#\{i \in \mathbb{Z}^d / \ \|i\| \leqslant p\}$ and $s_p = \#\{i \in \mathbb{Z}^d / \ \|i\| = p\}$ for $\|i\| = \max\{|i_1|, \ldots, |i_d|\}$ we obtain $b_p = (2p+1)^d$ and $s_p = b_p - b_{p-1} \leqslant Kp^{d-1}$ for a constant $K > 0$.

Proof of lemma 2.2. $\displaystyle\sum_{\|j\|>p} a_j = \sum_{q>p} \sum_{\|j\|=q} e^{-\beta q} \leqslant Kp^{d-1} \sum_{q>p} e^{-\beta q} = \mathcal{O}\left(p^{d-1} e^{-\beta p}\right)$. We thus find a constant C_1 such that

$$\delta_r \leq C_1 \inf_p \left\{p^{d-1} e^{-\beta p} + e^{-\alpha \frac{r}{p}}\right\} = C_1 \inf_p \left\{e^{-\beta p + (d-1)\ln p} + e^{-\alpha \frac{r}{p}}\right\}.$$

Assume $p \sim \sqrt{\alpha r / \beta}$, there is a constant C_2 such that : $\delta_r \leq C_2 r^{\frac{d-1}{2}} e^{-\sqrt{\alpha \beta r}}$. □

Proof of lemma 2.3. As before, $\sum_{\|j\|>p} a_j \leq K \frac{p^{d-\beta}}{\beta - d}$. Hence $\delta_r \leqslant c \cdot \inf_p \left\{ e^{-\alpha r/p} + \dfrac{p^{d-\beta}}{\beta - d} \right\}$. Choose $p \sim \frac{\alpha r}{(\beta - d) ln r}$, thus there exists some constant C_3 with $\delta_r \leq C_3 \left(\frac{r}{\ln r}\right)^{d-\beta}$. □

2.4.4 Proof of proposition 2.1

Models with finite interactions

We assume first that there exist $k \geqslant 1$ and $j_1, \ldots, j_k \in I$ such $F(x; u)$ only depends on $x_{j_1}, \ldots, x_{j_k}$ for each $x = (x_j)_{j \neq 0} \in E^I$. Hence writing a_i instead of a_{j_i} for $1 \leq i \leq k$, we have

$$\|F(x; u) - F(y; u')\| \leq \sum_{i=1}^k a_i \|x_{j_i} - y_{j_i}\| + b\|u - u'\|, \qquad a = \sum_{i=1}^k a_i < 1$$

Now $h : E^k \times E' \to E$ is such that $F(x; u) = h(x_{j_1}, \ldots, x_{j_k}, u)$. We will denote $\rho = \max\{\|j_1\|, \ldots, \|j_k\|\}$.

Lemma 2.10 *1) If the random field ξ is $\eta-$weakly dependent (the weak dependence coefficients are denoted $\eta_\xi(r)$) then X is $\eta-$weakly dependent with, for $C>0$,*

$$\eta(r) \leq C \inf_{n\in\mathbb{N}^*} \left\{a^n + k^n \eta_\xi\left((r-2\rho n)\vee 0\right)\right\}$$

2) If the random field ξ is $\lambda-$weakly dependent (the weak dependence coefficients are denoted $\lambda_\xi(r)$) then X is $\lambda-$weakly dependent, for $C>0$,

$$\lambda(r) \leq C \inf_{n\in\mathbb{N}^*} \left\{a^n + k^{2n} \lambda_\xi\left((r-2\rho n)\vee 0\right)\right\}$$

Proof of lemma 2.10. We will use the lemma 2.6 and the following useful lemma 2.11.

Lemma 2.11 *1. For every x and $y \in C^{(\mathbb{Z}^d)}$ we have $\|\Phi(0)(x) - \Phi(0)(y)\| \leq b\|x_0 - y_0\|$ and if $n \geq 2$,*

$$\|\Phi^{(n)}(0)(x) - \Phi^{(n)}(0)(y)\| \\ \leq \sum_{l=1}^{n-1} \sum_{i_1,\ldots,i_l=1}^{k} a_{i_1}\cdots a_{i_l} b \|x_{j_{i_1}+\cdots+j_{i_l}} - y_{j_{i_1}+\cdots+j_{i_l}}\| + b\|x_0 - y_0\|$$

2. Fix $x \in C^{(Z^d)}$. Then $\Phi(0)(x)$ only depends on x_0 and $\Phi(0)$ defines a $b-$Lipschitz function on C. We set $K_1 = b$ and $p_1 = 1$.
For $n \geqslant 2$ we set $A_n = \bigcup_{l=1}^{n-1}\{j_{i_1} + \cdots + j_{i_l} \,/\, 1 \leq i_1,\ldots,i_l \leq k\} \cup \{0\}$, $p_n = |A_n|$, the cardinal of A_n and $K_n = b\frac{1-a^n}{1-a}$. Then $\Phi^{(n)}(0)(x)$ only depends on x_j for $j \in A_n$. Moreover $\Phi^{(n)}(0)$ defines a Lipschitz function on C^{p_n} and Lip $\left(\Phi^{(n)}(0)\right) \leq K_n$.

Proof of lemma 2.11.
– The first point is easy to check. For $n \geq 2$ we use induction. For $n = 2$

$$\begin{aligned} \|\Phi^{(2)}(0)(x) - \Phi^{(2)}(0)(y)\| &\leq \sum_{i=1}^{k} a_i \|F(0,x_i) - F(0,y_i)\| + b\|x_0 - y_0\| \\ &\leq \sum_{i=1}^{k} a_i b \|x_i - y_i\| + b\|x_0 - y_0\| \end{aligned}$$

Assuming that the inequality holds for an integer $n \geq 2$, we estimate $\phi_{n,x,y} = \|\Phi^{(n+1)}(0)(x) -$

$\Phi^{(n+1)}(0)(y)\|$:

$$\begin{aligned}
\phi_{n,x,y} &\leq \sum_{i=1}^{k} a_i \|\Phi^{(n)}(0)(\theta_{j_i}x) - \Phi^{(n)}(0)(\theta_{j_i}y)\| + b\|x_0 - y_0\| \\
&\leq \sum_{i=1}^{k} a_i \Big(\sum_{l=1}^{n-1} \sum_{1\leq i_1,\ldots,i_l\leq k} a_{i_1}\cdots a_{i_l} b \|x_{j_i+j_{i_1}+\ldots+j_{i_l}} - y_{j_i+j_{i_1}+\ldots+j_{i_l}}\| \\
&\quad + b\|x_{j_i} - y_{j_i}\| \Big) + b\|x_0 - y_0\| \\
&= \sum_{l=1}^{n} \sum_{1\leq i_1,\ldots,i_l\leq k} a_{i_1}\cdots a_{i_l} b \|x_{j_{i_1}+\ldots+j_{i_l}} - y_{j_{i_1}+\ldots+j_{i_l}}\| + b\|x_0 - y_0\|
\end{aligned}$$

Hence inequality holds for $n+1$.

– The case $n = 1$ is easy to check. For the first point we use induction. For $n = 2$ the result is a consequence of :

$$\Phi^{(2)}(x)(0) = h\left(h(0, x_{j_1}), \ldots, h(0, x_{j_k}), x_0\right)$$

Suppose now the result true for an integer $n \geq 2$. Then the identity

$$\Phi^{(n+1)}(0)(x) = h\left(\Phi^{(n)}(0)(\theta_{j_1}x), \ldots, \Phi^{n}(0)(\theta_{j_k}x), x_0\right)$$

shows that $\Phi^{(n+1)}(0)(x)$ only depends of coordinates $(x_{j_i+j})_{1\leq i\leq k, j\in A_n}$ and x_0 that is to say coordinates $(x_j)_{j\in A_{n+1}}$.

For the second point, we use inequality in 1. We have :

$$\begin{aligned}
\phi_{n,x,y} &\leq \sum_{l=1}^{n-1} \sum_{1\leq i_1,\ldots,i_l\leq k} a_{i_1}\cdots a_{i_l} b \|x_{j_{i_1}+\ldots+j_{i_l}} - y_{j_{i_1}+\ldots+j_{i_l}}\| + b\|x_0 - y_0\| \\
&\leq b\Big(\sum_{l=1}^{n-1} a^l + 1\Big) \sum_{j\in A_n} \|x_j - y_j\| = b \cdot \frac{1-a^n}{1-a} \sum_{j\in A_n} \|x_j - y_j\|
\end{aligned}$$

End of the proof of lemma 2.10. We recall the notation $X_t^n = \Phi^{(n)}(0)\big((\xi_{t-j})_j\big)$ for $n \in \mathbb{N}^*$ and $t \in \mathbb{Z}^d$. Set $f_1 = f(X_{s_1}, \ldots, X_{s_u})$, $g_1 = g(X_{t_1}, \ldots, X_{t_v})$ and

$$f_1' = f\left(X_{s_1}^n, \ldots, X_{s_u}^n\right), \qquad g_1' = g\left(X_{t_1}^n, \ldots, X_{t_v}^n\right)$$

For each $t \in \mathbb{Z}^d$, if $n \in \mathbb{N} \setminus \{0,1\}$ then $A_{t,n} = \{t\} - A_n$. If $\|s_i - t_l\| \geq r$ for $1 \leq i \leq u$ and $1 \leq l \leq v$ then $d(A_{s_i,n}, A_{t_l,n}) \geq (r - 2\rho n) \vee 0 = d_{r,n}$. Thus

$$\begin{aligned}
|\mathrm{Cov}(f_1, g_1)| &\leq |\mathrm{Cov}(f_1 - f_1', g_1)| + |\mathrm{Cov}(f_1', g_1 - g_1')| + |\mathrm{Cov}(f_1', g_1')| \\
&\leq 4\mathbb{E}|f_1 - f_1'| + 4\mathbb{E}|g_1 - g_1'| \\
&\quad + \psi(up_n, vp_n, K_n \mathrm{Lip}\,(f), K_n \mathrm{Lip}\,(g))\varepsilon_\xi(d_{r,n}) \\
&\leq (4\mathrm{Lip}\,(f)u + 4\mathrm{Lip}\,(g)v)a^n \|X_0\|_1 \\
&\quad + \psi(up_n, vp_n, K_n \mathrm{Lip}\,(f), K_n \mathrm{Lip}\,(g))\varepsilon_\xi(d_{r,n})
\end{aligned}$$

Note that this result is still true for $n = 1$.

1) Under $\eta-$weak dependence, $\psi(u, v, a, b) = au + bv$,

$$|\mathrm{Cov}(f_1, g_1)| \leq (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g)(4a^n \|X_0\|_1 + K_n p_n \eta_\xi(d_{r,n})$$

Thus $|\mathrm{Cov}(f_1, g_1)| \leq (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g)\eta(r)$ where

$$\eta(r) \leq \inf_{n \in \mathbb{N}^*} \{4a^n \|X_0\|_1 + K_n p_n \eta_\xi(d_{r,n})\}$$

2) With $\lambda-$weak dependence $\psi(u, v, a, b) = au + bv + abuv$,

$$|\mathrm{Cov}(f_1, g_1)| \leq (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g + uv\mathrm{Lip}\, f\mathrm{Lip}\, g)(4a^n \|X_0\|_1 + K_n p_n \lambda_\xi(d_{r,n})$$

Now $|\mathrm{Cov}(f_1, g_1)| \leq (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g + uv\mathrm{Lip}\, f\mathrm{Lip}\, g)\lambda(r)$ with

$$\lambda(r) \leq \inf_{n \in \mathbb{N}^*} \{4a^n \|X_0\|_1 + K_n p_n^2 \lambda_\xi(d_{r,n})\}$$

As $(K_n)_n$ is bounded and $p_n \leq \sum_{l=1}^{n-1} k^l = \frac{k - k^n}{1 - k}$ for $n \geq 2$, we obtain the proposed bounds.

We now prove that $\lim_{r \to \infty} \lambda(r) = 0$. We suppose that the sequence $(\lambda_\xi(r))_r$ nonincreasing without loss of generality. We use the bound

$$\lambda(r) \leq C \inf_{N + 2\rho n = r, n \in \mathbb{N}^*} \left\{a^n + k^{2n} \lambda_\xi(N)\right\}$$

If $N \in \mathbb{N}$, we choose $n_N = [log(\lambda_\xi(N))/(\log a - 2\log k)]$. Note that $\lim_{N \to \infty} n_N = \infty$ and $\lim_{N \to \infty} \big(a^{n_N} + k^{2n_N} \lambda_\xi(N)\big) = 0$. For $r \geq r_N = N + 2\rho n_N$, we have $N + r - r_N + 2\rho n_N = r$, hence :

$$\lambda(r) \leq a^{n_N} + k^{2n_N} \lambda_\xi(N + r - r_N) \leq a^{n_N} + k^{2n_N} \lambda_\xi(N) \to_{N \to \infty} 0$$

Hence $\lim_{r \to \infty} \lambda(r) = 0$. Analogously, $\lim_{r \to \infty} \eta(r) = 0$. $\square$

2.4.5 General case

Recall that we have denoted $(X_{p,t})_{t\in\mathbb{Z}^d}$ for $s>0$ the unique solution of the equation $Z_t = F\left(\left(Z_{t-j}1_{0<\|j\|\le p\}}\right)_{j\neq 0};\xi_t\right)$. Denote $f_1 = f(X_{s_1},\dots,X_{s_u})$, $g_1 = g(X_{t_1},\dots,X_{t_v})$, $f_1' = f(X_{p,s_1},\dots,X_{p,s_u})$ and $g_1' = f(X_{p,t_1},\dots,X_{p,t_v})$, then

$$\begin{aligned}
|\mathrm{Cov}(f_1,g_1)| &\le |\mathrm{Cov}(f_1-f_1',g_1)| + |\mathrm{Cov}(f_1',g_1-g_1')| + |\mathrm{Cov}(f_1',g_1')| \\
&\le 4\,\|X_0 - X_{s,0}\|_1(u\mathrm{Lip}\,f + v\mathrm{Lip}\,g) + |\mathrm{Cov}(f_1',g_1')|
\end{aligned}$$

Recall that from the proof of lemma 2.8 we have : $\|X_{p,0}-X_0\|_1 \le \dfrac{\|X_0\|_1}{1-a}\sum_{\|j\|>p} a_j$. Moreover, the field $X_{p,t}$ is $k-$dependent with $k=(2p)^d$.

– Suppose first that the random field ξ is $\eta-$weakly dependent. From proposition 2.10,

$$|\mathrm{Cov}(f_1',g_1')| \le (u\mathrm{Lip}\,f + v\mathrm{Lip}\,g)C\inf_{n\in\mathbb{N}^*}\left\{a^n + p^{dn}\eta_\xi\left((r-2pn)\vee 0\right)\right\}$$

for a suitable positive constant C.
Hence we bound $|\mathrm{Cov}(f_1,g_1)|$ by,

$$(u\mathrm{Lip}\,f + v\mathrm{Lip}\,g)C\left(\sum_{\|j\|>p} a_j + \inf_{n\in\mathbb{N}^*}\left\{a^n + p^{dn}\eta_\xi\left((r-2\rho n)\vee 0\right)\right\}\right)$$

for another positive constant denoted C. Then we obtain the proposed bound.

– Suppose that the random field ξ is $\lambda-$weakly dependent. From proposition 2.10, $|\mathrm{Cov}(f_1',g_1')|$ is bounded by

$$(u\mathrm{Lip}\,f + v\mathrm{Lip}\,g + uv\mathrm{Lip}\,f\mathrm{Lip}\,g)\inf_{n\in\mathbb{N}^*}\left\{a^n + p^{2dn}\lambda_\xi\left((r-2pn)\vee 0\right)\right\}$$

up a suitable positive constant C. Hence we bound $|\mathrm{Cov}(f_1,g_1)|$ by,

$$\begin{aligned}
&(u\mathrm{Lip}\,f + v\mathrm{Lip}\,g + uv\mathrm{Lip}\,f\mathrm{Lip}\,g) \\
&\qquad\qquad \times\left(\sum_{\|j\|>p} a_j + \inf_{n\in\mathbb{N}^*}\left\{a^n + p^{2dn}\lambda_\xi\left((r-2pn)\vee 0\right)\right\}\right)
\end{aligned}$$

up to another positive constant C. Then we obtain the proposed bound.

2.4.6 Results on causality

Proof of proposition 2.2

We will use here the Euclidean norm on $\mathbb{R}^d$. We proceed by induction on d.
For $d=1$, if there exists $r_1,r_2\in B$ such that $r_1>0$ and $r_2<0$ then $B\cap(-B)\neq\{0\}$. Then we can

choose $b_1 = 1$ if $B \subset \mathbb{R}_+$ or $b_1 = -1$ if $B \subset \mathbb{R}_-$.
Suppose the result true for $d-1$. We first define b_1.
1) If B° is empty, since B is convex and contain 0 there exists $b_1 \in \mathbb{R}^d \setminus \{0\}$ such that $B \subset H = \{x \in \mathbb{R}^d / x.b_1 = 0\}$($\cdot$ denotes the scalar product in $\mathbb{R}^d$).
2) Now if B° is not empty, like $B \cap (-B) = \{0\}$ it is clear that $0 \notin B^\circ$. Moreover B° is still convex and by application of the Hahn-Banach theorem ([2], theorem 3.3, page 108), there exists $b_1 \in R^d \setminus \{0\}$ such that $B^\circ \subset \{x \in \mathbb{R}^d / x.b_1 \geq 0\}$. Like for a convex $\overline{B^\circ} = \overline{B}$, then the same inclusion holds for B. We set here $H = \{x \in \mathbb{R}^d / x.b_1 = 0\}$.
We consider now the convex cone $C = B \cap H$. If g denote an isomorphism between H and $\mathbb{R}^{d-1}$, then $g(C)$ is a convex cone of $\mathbb{R}^{d-1}$ such that $g(C) \cap \big(-g(C)\big) = \{0\}$. Hence there exists a basis $c = (c_2, \dots, c_d)$ such that $g(C) \subset \{x \in \mathbb{R}^{d-1} / 0 \leq_c x\}$. For $i = 2, \dots, d$ we set $b_i = g^{-1}(c_i)$. Then $b = (b_1, \dots, b_d)$ is a basis of $\mathbb{R}^d$ and if $x = x_1 b_1 + \ldots + x_d b_d \in B$, we have by the preceding two points $x_1 \geq 0$. Suppose that $x_1 = 0$, then $x \in C$ and $g(x) \geq_c 0 \Rightarrow (x_2, \dots, x_d) \geq_{lex} 0$ in $\mathbb{R}^{d-1}$. Hence $(x_1, \dots, x_d) \geq_{lex} 0$ in $\mathbb{R}^d$, in other word $x \geq_b 0$.

Proof of lemma 2.5

Proof of lemma 2.5 Denote for $p \in \mathbb{N}^*$, $Y_p = F\big((X_j \mathbb{1}_{0<\|j\|\leq p})_j; \xi_0\big)$.
1) We first prove that for $p \in \mathbb{N}^*$, $Y_p \in \mathbb{L}^m$.
Recall that here F is measurable wrt $\mathfrak{F}_{\widetilde{A}} \otimes \mathcal{B}(E')$. Let $x \in E^I$, then using the independence between ξ_0 and $\sigma\big(X_j, j \in A\big)$ and the condition **(H3)**, we have :

$$\begin{aligned}\mathbb{E}\big(\|Y_p - F(0;\xi_0)\|^m / X_j = x_j, j \in A\big) &= \mathbb{E}\left\|F((x_j \mathbb{1}_{0<\|j\|\leq p})_j;\xi_0) - F(0;\xi_0)\right\|^m \\ &\leq \left(\sum_{0<\|j\|\leq p} a_j \|x_j\| \right)^m\end{aligned}$$

Hence by integration :

$$\|Y_p - F(0;\xi_0)\|_m \leq \Big\| \sum_{0<\|j\|\leq p} a_j \|X_j\| \Big\|_m \leq \|X_0\|_m$$

As $F(0;\xi_0) \in \mathbb{L}^m$, we obtain the result.
It is enough to prove that $(Y_p)_{p\in\mathbb{N}^*}$ is a Cauchy sequence in $\mathbb{L}^m$. Using the same method as in 1), we obtain if $q > 0$:

$$\|Y_{p+q} - Y_p\|_m \leq \|X_0\|_m \sum_{\|j\|>p} a_j$$

This inegality imply the result.
2) Using the same method as in 1), we have for $p \in \mathbb{N}^*$:

$$\left\|Y_p - Y'_p\right\|_m \leq \sum_{0<\|j\|\leq p} a_j \left\|Y_j - Y'_j\right\|_m$$

Hence the result follows with $p \to \infty$. □

2.4.7 Proof of theorem 2.3

Existence

If $H \in \mathbb{L}^m_{\widetilde{A}}(\mu)$, we denote by Y the random field defined as $Y_j = H \circ \theta_j(\xi)$ for $j \in \mathbb{Z}^d$. If $j \in A$, then $H \circ \theta_j$ is measurable wrt $\sigma(\pi'_{j+j'}/j' \in \widetilde{A}) \subset \mathfrak{F}'_{\widetilde{A}}$. Hence if $p \in \mathbb{N}^*$, $\Phi_p(H) \in \mathbb{L}^m_{\widetilde{A}}(\mu)$ and by the lemma 6.5, ξ_0 is independant of $\sigma(Y_j/j \in A)$. By application of the lemma 2.5, $\Phi_p(H)$ converges to an element of $\mathbb{L}^m_{\widetilde{A}}(\mu)$ witch is $F\big((H \circ \theta_j)_{j\neq 0}; \pi_0\big)$.
Lets show that the application $\Phi : \mathbb{L}^m_{\widetilde{A}}(\mu) \mapsto \mathbb{L}^m_{\widetilde{A}}(\mu)$ defined as $\Phi(H) = F\big((H \circ \theta_j)_{j\neq 0}; \pi_0\big)$ is contraction in $\mathbb{L}^m_{\widetilde{A}}(\mu)$. If $H, H' \in \mathbb{L}^m_{\widetilde{A}}(\mu)$ then the two random fields Y and Y' defined as $Y_j = H \circ \theta_j(\xi)$ and $Y'_j = H' \circ \theta_j(\xi)$ for $j \in \mathbb{Z}^d$ verify the assumptions of lemma 2.5. Indeed $\sigma(Y_j, Y'_j/j \in A) \subset \sigma(\xi_{j+j'}/j \in A, j' \in \widetilde{A})$ and using the lemma 6.5 we deduce the independence between ξ_0 and $\sigma(Y_j, Y'_j/j \in A)$. Hence, we have :

$$\begin{aligned}\left\|\Phi(H)(\xi) - \Phi(H')(\xi)\right\|_m &\leq \sum_{j\in A} a_j \left\|H \circ \theta_j(\xi) - H' \circ \theta_j(\xi)\right\|_m \\ &= \sum_{j\in A} a_j \left\|H(\xi) - H'(\xi)\right\|_m\end{aligned}$$

witch shows the result.
The construction of X_t comes from theorem 2.2. The variable $H(\xi)$ being measurable wrt $\sigma\left(\xi_j; j \in \widetilde{A}\right)$ measurability of X_t is simply deduced. Then unicity is a consequence of the application of the fixed point theorem. □

Weak dependence

Weak dependence of the solution is as in § 2.4.3 where **(H2)** replaces **(H3')**. The case of finite range corresponds to k-Markov systems on a finite causal set. To prove lemma 2.7, we use **(H3')** and independence of rvs $\left(X_{t-j_1}, \ldots, X_{t-j_k}, X^{(r)}_{t-j_1}, \ldots, X^{(r)}_{t-j_k}\right)$ and ξ_t to derive (2.12). In the general case we note $(X_{p,t})_t$ the solution of $Z_t = F\left((Z_{t-j}1_{\{j\in A_p\}})_j; \xi_t\right)$ with $A_p = \{t \in A/\|t\| \leqslant p\}$ and we conclude as in lemma 2.9. □

2.4.8 Proof of lemma 2.6

1) From the fixed point theorem, we deduce that for each $\varepsilon > 0$:

$$\begin{aligned}\sum_{n\geq 1}\mathbb{P}\left(\|X_t-\Phi^{(n)}(0)(\xi_{t-j},j\in\mathbb{Z}^d)\|\geq\varepsilon\right) &\leq \frac{1}{\varepsilon}\sum_{n\geq 1}\|X_t-\Phi^{(n)}(0)(\xi_{t-j},j\in\mathbb{Z}^d)\|_1\\ &< \infty\end{aligned}$$

Hence by the Borel-Cantelli lemma, we deduce $\lim_{n\to\infty}\Phi^{(n)}(0)\left(\xi_{t-j},j\in\mathbb{Z}^d\right)=X_t$ *a.s.*
2) We use induction. For $n=1$

$$\begin{aligned}\|X_t-\Phi_p(0)(\xi_{t-j},j\in\mathbb{Z}^d)\|_1 &= \|F\left((X_{t-j})_j,\xi_t\right)-F(0,\xi_t)\|_1\\ &\leq a\|X_0\|_1\\ &\leq a\|X_0\|_1+\|X_0\|_1\sum_{\|j\|>p}a_j\end{aligned}$$

Suppose the result true for an integer $n\geq 1$, then

$$\begin{aligned}\|X_t-X_{p,t}^{n+1}\|_1 &\leq \sum_{\|k\|\leq p}a_k\Big\|X_{t-k}-X_{p,t-k}^n\|_1+\|X_0\|_1\sum_{\|k\|>p}a_k\\ &\leq a\Big(a^n\|X_0\|_1+\frac{1-a^n}{1-a}\|X_0\|_1\sum_{\|k\|>p}a_k\Big)+\|X_0\|_1\sum_{\|k\|>p}a_k\\ &= a^{n+1}\|X_0\|_1+\frac{1-a^{n+1}}{1-a}\|X_0\|_1\sum_{\|k\|>p}a_k\end{aligned}$$

2.4.9 Proofs for the section 2.3

Proof of corollary 2.1

Here $F(x;u)=f(x_{\ell_1},\ldots,x_{\ell_k})+h(u)g(x_{\ell_1},\ldots,x_{\ell_k})+u$. Condition **(H1)** is easy to check and *e.g.* in the first case,

$$\|F(z;\xi_0)-F(z';\xi_0)\|_m\leq\sum_{i=1}^k b_i\|z_{\ell_i}-z'_{\ell_i}\|+\|\gamma_0\|_\infty\sum_{i=1}^k c_i\|z_{\ell_i}-z'_{\ell_i}\|.$$

For dependent inputs, we remark that $(z,u)\mapsto F(z;u)$ is a Lipschitz function in order to apply proposition 2.1.$\square$

Proof of proposition 2.3

Normal convergence in $\mathbb{L}^m$ will justify all the forthcoming manipulations of series. We only consider the more complicated causal case. In order to prove that $X_t\in\mathbb{L}^m$ we will prove the normal convergence of the series.

Set $S = \|\xi_t\| + \sum_{i=1}^{+\infty} \sum_{j_1,\dots,j_i\in A} \|\alpha_t^{j_1}\cdots\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\xi_{t-(j_1+\dots+j_i)}\|_m$, we notice from causality that indices t and $(t-(j_1+\cdots+j_\ell))$ are distinct if $1\leqslant \ell \leqslant i$ hence the independence of inputs implies

$$\|\alpha_t^{j_1}\cdots\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\xi_{t-(j_1+\dots+j_i)}\|_m \leq= \|\alpha_t^{j_1}\|_m\cdots\|\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\|_m\|\xi_{t-(j_1+\dots+j_i)}\|_m$$

$$\begin{aligned} S &\leq \|\zeta_t\|_m + \sum_{j\in A}\sum_{j_1,\dots,j_i\in A} \|\alpha_t^{j_1}\|_m\cdots\|\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\|_m\|\zeta_{t-(j_1+\dots+j_i)}\|_m \\ &= \|\zeta_0\|_m\left(1+\frac{b}{1-b}\right) < \infty. \end{aligned}$$

In order to prove that X_t is solution of the equation, we expand it :

$$\begin{aligned} X_t &= \zeta_t + \sum_{j_1\in A}\alpha_t^{j_1}\ \zeta_{t-j_1} + \sum_{i=2}^{\infty}\sum_{j_1,\dots,j_i\in A}\alpha_t^{j_1}\cdots\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\zeta_{t-(j_1+\dots+j_i)} \\ &= \zeta_t + \sum_{j_1\in A}\alpha_t^{j_1}\left(\zeta_{t-j_1} + \sum_{i=2}^{\infty}\sum_{j_2,\dots,j_i\in A}\alpha_{t-j_1}^{j_2}\cdots\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\zeta_{t-j_1-(j_2+\dots+j_i)}\right) \\ &= \zeta_t + \sum_{j_1\in A}\alpha_t^{j_1}X_{t-j_1} \end{aligned}$$

Here $F(x;(u,v)) = \sum_{j\in A} u_jx_j + v$ and we use notations in **(H3)**. As ξ is iid, the variables $(Z(\xi), Z'(\xi))$ are $(\alpha_0^j)_{j\in A}$ are independent and

$$\|F(z;\zeta_0) - F(z';\zeta_0)\|_m \leq \sum_{j\in A}\|\alpha_0^j\|_m\|z_j - z'_j\|$$

Since $b = \sum_{j\in A}\|a_0^j\|_m < 1$, **(H3)** holds.
In the first non-causal case the above inequalities are only changed by using the bound

$$\|\alpha_t^{j_1}\cdots\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\xi_{t-(j_1+\dots+j_i)}\|_m \leqslant \|\alpha_t^{j_1}\|_\infty\cdots\|\alpha_{t-j_1-\dots-j_{i-1}}^{j_i}\|_\infty\|\xi_{t-(j_1+\dots+j_i)}\|_m. \ \square$$

Proof of proposition 2.4

Here **(H1)** holds and with the notation in **(H3)** :

$$\|F(z,\xi_0) - F(z',\xi_0)\|_m \leq \sum_{j\in\mathbb{Z}^d\setminus\{0\}}\|\alpha_j\|\|\xi_0\|_m\|z_j - z'_j\|.$$

The proposed solution is in $\mathbb{L}^m$ from normal convergence of series

$$\begin{aligned} \|X_t\|_m &\leq \|\xi_t\|_m\Big(\|a\| + \sum_{k=1}^{\infty}\sum_{j_1,\dots,j_k\in A}\|\alpha_{j_1}\|\|\xi_{t-j_1}\|_m\cdots\|\alpha_{j_k}\|\|\xi_{t-j_1-\dots-j_k}\|_m\|a\|\Big) \\ &= \|\xi_0\|_m\|a\|\Big(1+\frac{b\|\xi_0\|_m}{1-b\|\xi_0\|_m}\Big) < \infty. \end{aligned}$$

Substitutions prove that this process is a solution of the equation.

$$\begin{aligned}
X_t &= \xi_t\Big(a + \sum_{k=1}^{\infty} \sum_{j_1,\dots,j_k \in A} \alpha_{j_1}\xi_{t-j_1} \cdots \alpha_{j_k}\xi_{t-j_1-\dots-j_k} a\Big) \\
&= \xi_t\Big(a + \sum_{j_1 \in A} \alpha_{j_1}\xi_{t-j_1}\Big(a + \sum_{k=2}^{+\infty} \alpha_{j_2}\xi_{t-j_1-j_2} \dots \alpha_{j_k}\xi_{t-j_1-j_2-\dots-j_k}\Big)\Big) \\
&= \xi_t\Big(a + \sum_{j_1 \in A} \alpha_{j_1} X_{t-j_1}\Big). \qquad \square
\end{aligned}$$

Bibliographie

[1] Comets, F., Fernandez, R., Ferrari, P. A. (2002) *Processes with Long Memory : Regenerative Construction and Perfect Simulation.* Ann. Appl. Probab. vol. 12-3, 921-943.

[2] Conway, J. B. (1990) *A course in functional analysis.* Graduate Texts in Mathematics, Vol. 96 (2nd edition). Springer, New-York.

[3] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., Prieur, C. (2007) *Weak dependence, examples and applications,* monograph, to appear as a Lecture Notes in Statistics. Springer Verlag, New-York.

[4] Dobrushin, R. L. (1970) *Prescribing a system of random variables by conditional distributions.* Th. Probab. Appl. 15, 458-486.

[5] Doukhan, P., Latour, A., Truquet, L., Wintenberger, O. (2007) *Integer valued random fields with infinite interactions.* In preparation.

[6] Doukhan, P., Louhichi, S. (1999) *A new weak dependence condition and applications to moment inequalities.* Stoch. Proc. Appl. 84, 313-342.

[7] Doukhan, P., Teyssière, G. and Winant, P. (2006) *Vector valued ARCH(∞) processes,* in *Dependence in Probability and Statistics*, Patrice Bertail, Paul Doukhan, and Philippe Soulier (Editors.) Lecture Notes in Statistics, Springer, New York.

[8] Doukhan, P., Wintenberger, O. (2006) *Weakly dependent chains with infinite memory* Preprint available in `http ://samos.univ-paris1.fr/ppub2005.html`.

[9] Föllmer, H. (1988) *Random Fields and Diffusion Processes.* École d'été de Probabilités de St Flour, Lecture Notes in Math. 1362, 101-204.

[10] Guyon, X. (1995) *Random fields on a network : modeling, statistics, and applications.* Graduate Texts in Mathematics. Springer-Verlag, New-York.

[11] Georgii, H. O. (1988) *Gibbs measures and phase transition.* Studies in Mathematics 9, De Gruyter.

[12] Helson, H., Lowdenslager, D. (1959) *Prediction theory and Fourier series in several variables.* Acta mathematica, 99, 165-202.

[13] Loubaton, P. (1989) *Champs stationnaires au sens large sur $\mathbb{Z}^2$: Propriétés structurelles et modèles paramétriques.* Traitement du signal, 6-4, 223-247.

[14] Shashkin, A. P. (2005) *A weak dependence property of a spin system*, Preprint.

[15] Tjostheim, D. (1986) *Some doubly stochastic time series models.* J. Time Series Anal. 7, 51-72.

Chapitre 3

Weak Dependence, Models and Some Applications

Abstract

The paper is devoted to recall weak dependence conditions from Dedecker *et al.* (2007)'s monograph ; the main basic results are recalled here and we go further in some new applications. We develop here several models of weakly dependent processes and random fields. Among them an ARCH(∞) model is considered with statistical applications to ordinary least squares. A last part aims at proving new asymptotic results for weakly dependent random fields. Such applications are indeed the main proof of the interest of this theoretical notion which measures the asymptotic decorrelation of a process.

Note

The content of this part is based on a paper, written in collaboration with Paul Doukhan and Nathanaël Mayo.

3.1 Introduction

The present work aims at answering the delicate question *"how to weaken independence ?"* up to really useful statistical applications. A first answer to this problem was the mixing assumption introduced by Rosenblatt in 1956 (see [13]), however a simple example of an AR(1) model with Bernoulli innovation is proved in Andrews (1984) [1] to be nonmixing. Extending Andrews's ideas we provide here a reader with a new and heteroskedastic ARCH(1) nonmixing model. The idea of weak dependence developed in the recent monograph Dedecker *et al.* (2007), [11] gives a reasonable answer to the previous question.
Different conditions of weak dependence are thus introduced and compared here. Further, a large

amount of models are detailed here and proved to be weakly dependent, among which figure the previous counterexamples, see also [14]. *E.g.* some recent extensions of ARCH models are weakly dependent, among them ARCH(∞) models are considered and we develop a specific parametric estimation procedure, *ordinary least squares* (OLS). Our results are proved here for a *toy model* but nonparametric procedures should avoid such question. A final section is devoted to new asymptotic results for random fields and mimicking Bulinskii & Shashkin (2006), [9], we obain a moment inequality, a central limit theorem and a functional central limit theorem for the partial sums process of a weakly dependent random field. For this, those authors make use of a nice Marcinkiewicz-Zygmund inequality with order > 2. From this moment inequality we first derive a Central Limit Theorem through Stein's technique and also a tightness argument.

The paper is organized as follows. A first section is devoted to state the problem and to recall the mixing condition as well as to state some counterexamples in order to introduce weak dependence conditions. We also recall some basic inequalities related to those weak dependence conditions. We then list some examples of weakly dependent models. The third section details ordinary least squares estimation procedure and its asymptotic properties for the special case of ARCH(∞) models. A last section proves a Donsker type result for the partial sums process of a weakly dependent random field.

3.2 Weak dependence

3.2.1 Independence

Consider the $\sigma-$algebras, $\sigma(P)$ and $\sigma(F)$ generated by random variables P and F, independence of such random variables only writes as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \qquad \forall A \in \sigma(P),\ \forall B \in \sigma(F).$$

Independence also writes $\mathrm{Cov}(f(P), g(F)) = 0$ for all f, g with $\|f\|_\infty, \|g\|_\infty \leqslant 1$. Later on, the variables P and F will be denoted Past and Future for both the times series case and the random field setting. For a process $X = (X_t)_{t \in \mathbb{T}}$ we set

$$P = (X_{s_1}, \dots, X_{s_u}), \qquad F = (X_{t_1}, \dots, X_{t_v})$$

Since no phenomena are really independent from each others, a first question is here, how to weaken those relations.

3.2.2 Mixing

Assume that $\mathbb{T} = \mathbb{Z}$ and $s_1 \leqslant \cdots \leqslant s_u$, $t_1 \leqslant \cdots \leqslant t_v$, with $r = t_1 - s_u$ large, a first answer to our question is to set

$$\alpha(\sigma(P), \sigma(F)) = \sup_{\|f\|_\infty, \|g\|_\infty \leqslant 1} |\mathrm{Cov}(f(P), g(F))| \,.$$

Rosenblatt (1956)'s mixing coefficient is

$$\alpha(r) = \sup_{u,v} \max_{\substack{s_1 \leqslant \cdots \leqslant s_u \\ t_1 \leqslant \cdots \leqslant t_v \\ r = t_1 - s_u}} \alpha(\sigma(P), \sigma(F))$$

The process X is strong mixing if $\alpha(r) \downarrow 0$ as $r \uparrow \infty$. A nice theory was developed in this mixing setting case ; see Bradley (2007) [6], Doukhan (1994) [13], and Rio (2000) [27], also for the references herein. Andrews (1984) [1]'s simple example is however not mixing

$$X_t = \frac{1}{2}\left(X_{t-1} + \xi_t\right), \qquad \xi_t \sim b\left(\frac{1}{2}\right), \text{ iid.}$$

In his paper, Andrews (1984) [1] gives a proof adapted here to the following LARCH model :

$$X_t = \xi_t(1 + aX_{t-1}) \tag{3.1}$$

where $\mathbb{P}(\xi_0 = 1) = \mathbb{P}(\xi_0 = -1) = 1/2$. It is well know that $a < 1$ is a sufficient condition for existence and uniqueness in $\mathbb{L}^m$, $m \geq 1$, of a stationary solution of equation (3.1) (see [7]), moreover the solution writes : $X_t = \xi_t + \sum_{j \geq 1} a^j \xi_t \cdots \xi_{t-j}$.

Proposition 3.1 *We suppose* $a \in \left(\frac{3-\sqrt{5}}{2}, 1/2\right]$. *Then the stationary solution of equation (3.1) is not strongly mixing.*

Proof. We use the decomposition :

$$X_t = A_{t,n} + a^{n+1}\xi_t \cdots \xi_{t-n} X_{t-(n+1)}, \quad A_{t,n} = \xi_t + a\xi_t\xi_{t-1} + \ldots + a^n \xi_t \cdots \xi_{t-n} \tag{3.2}$$

for $n \in \mathbb{N}^*$. We prove as in Andrews (1984) that $\mathbb{P}\left(X_t \in A \big/ X_{t-(n+1)} \in B\right) = 1$, $(\forall n)$ $\mathbb{P}(X_{t-(n+1)} \in B) \neq 0$ and $\mathbb{P}\left(X_t \in A\right) < 1$, for some well chosen subsets A, B of $\mathbb{R}$. Set $U_t = (X_t \in A)$ and $V_{t-n-1} = (X_{t-(n+1)} \in B)$ then $\mathbb{P}(U_t \cap V_{t-n-1}) = \mathbb{P}(V_{t-n-1})$ and we derive from stationarity that $\mathbb{P}(V_{t-n-1}) = \mathbb{P}(V_0) \neq 0$ and $\mathbb{P}(U_t) = \mathbb{P}(U_0) < 1$; thus

$$\alpha_n \geqslant \mathbb{P}(U_t \cap V_{t-n-1}) - \mathbb{P}(U_t)\mathbb{P}(V_{t-n-1}) \geqslant \mathbb{P}(V_0)(1 - \mathbb{P}(U_0)) > 0.$$

The proof follows from three steps.

1. The values of the random variable $A_{t,n}$ are spaced of at least $2a^n$. Indeed two distinct values of $A_{t,n}$ are always spaced by a number $d = 2\sum_{i=0}^n \varepsilon_i a^i$ where for $i = 0, \ldots, n$, $\varepsilon_i \in \{-1, 0, 1\}$. As $l = \min\{i / 0 \leqslant i \leqslant n, \varepsilon_i \neq 0\}$ exists and $\varepsilon_l = 1$, we have $d = 2a^n$ if $l = n$ and if $l \leq n-1$ standard calculations yield

$$\begin{aligned} d \geq 2\Big(a^l - \sum_{i=l+1}^n a^i\Big) &\geq \frac{2}{1-a}\big(a^l - 2a^{l+1} + a^{n+1}\big) \\ &\geq \frac{2}{1-a}\big(a^l(1-2a) + a^{n+1}\big) \geq \frac{2}{1-a}\big(a^n(1-2a) + a^{n+1}\big) = 2a^n. \end{aligned}$$

2. We have $\mathbb{P}(a < |X_t| \leq 2) > 0$. Now $X_t \geq 1 + a - \sum_{i\geq 2} a^i > a$ for $a \in (0, 1/2]$ if $\xi_t = \xi_{t-1} = 1$. Moreover as $X_t \leq 1/(1-a) \leq 2$ for $a \in (0, 1/2]$ we conclude that $\mathbb{P}(a < |X_t| \leq 2) \geq 1/4$.
3. For $B = (-a, a)$ we have $\mathbb{P}(X_t \in B) > 0$ (by stationarity this expression does not depend on t). For this, observe first that $a \in \left]\frac{3-\sqrt{5}}{2}, 1/2\right]$ implies $1 - a - a^2 - a^3 - \cdots < a$; thus for $n_0 \geq 2$ large enough we get $1 - a - \cdots - a^{n_0} + \sum_{k\geq n_0+1} a^k < a$.
If $\xi_{t-i} = 1$ for $i \neq 1$ with $0 \leq i \leq n_0$, and $\xi_{t-1} = -1$, we have

$$0 \leq X_t \leq 1 - a - \cdots - a^{n_0} + \sum_{k\geq n_0+1} a^k < a.$$

Thus $\mathbb{P}\left(|X_t| < a\right) \geq 2^{-n_0-1}$. Now if $w_1, \ldots, w_k$ denote the values of $A_{t,n}$, we set $A = \bigcup_{i=1}^k]w_i - a^{n+2}, w_i + a^{n+2}[$. Using the decomposition (3.2) we infer that $X_t \in A$ if $|X_{t-(n+1)}| < a$ thus $\mathbb{P}\left(X_t \in A / X_{t-(n+1)} \in B\right) = 1$.
We prove here that $\mathbb{P}(X_t \in A) < 1$. If $a < |X_{t-(n+1)}| \leq 2$, then X_t writes as $w_i + c$ with $2a^{n+1} \geq |c| > a^{n+2}$. In this case $X_t \notin A$. Indeed $|X_t - w_i| > a^{n+2}$ and if, for example $c > 0$, we use point 1 and the fact that $a \leq 1/2$ to derive : $X_t < w_i + 2a^{n+1} \leq w_{i+1} - a^{n+2}$ provided w_{i+1} exists (else we have obviously $X_t \notin A$). And we obtain $X_t \notin A$ if $c > 0$. It is also the case if $c < 0$ with a similar argument. The result follows from $\mathbb{P}(X_t \in A) = \mathbb{P}(X_t \in A \cap |X_{t-(n+1)}| \leq a) \leq \mathbb{P}(|X_0| \leq a) < 1$. Moreover it is clear that $\mathbb{P}(|X_{t-(n+1)}| \leq a) \neq 0$. $\square$

3.2.3 Definition

We aim at defining weak dependence coefficients as sequences decaying to 0 and such that

$$|\mathrm{Cov}(f(P), g(F))| \leqslant \psi(u, v, \mathrm{Lip}\, f, \mathrm{Lip}\, g)\varepsilon(r)$$

Let us thus consider a process $(X_t)_{t\in\mathbb{Z}}$ with values in a Banach space $(E, \|\cdot\|)$. For $h : E^u \to \mathbb{R}$ $(u \in \mathbb{N}^*)$ we define

$$\mathrm{Lip}\, h = \sup_{(y_1,\ldots,y_u)\neq(x_1,\ldots,x_u)\in E^u} \frac{|h(y_1, \ldots, y_u) - h(x_1, \ldots, x_u)|}{\|y_1 - x_1\| + \cdots + \|y_u - x_u\|}.$$

Definition 3.1 [16] *A process $X = (X_n)_{n\in\mathbb{Z}}$ with values in $\mathbb{R}^d$ is (ε, Ψ)-weakly dependent process if, for some classes of functions $E^u, E^v \to \mathbb{R}$, $\mathcal{F}_u, \mathcal{G}_v$:*

$$\varepsilon(r) = \sup_{u,v} \sup_{\substack{s_1 \leqslant \cdots \leqslant s_u \\ t_1 \leqslant \cdots \leqslant t_v \\ r = t_1 - s_u}} \sup_{f\in\mathcal{F}_u, g\in\mathcal{G}_v} \frac{\left|Cov\Big(f(X_{s_1}, \ldots, X_{s_u}), g(X_{t_1}, \ldots, X_{t_v})\Big)\right|}{\Psi(f, g)} \to_{r\to\infty} 0.$$

Assume from now on that the classes of functions contain functions bounded by 1. Distinct functions Ψ yield η, θ, κ and λ weak dependence coefficients :

$$
\begin{array}{rclll}
\Psi(f,g) & = & u Lip\, f + v Lip\, g, & \textit{then denote} & \varepsilon(r) = \eta(r), \\
 & = & v Lip\, g, & \textit{then denote} & \varepsilon(r) = \theta(r), \\
 & = & uv Lip\, f \cdot Lip\, g, & \textit{then denote} & \varepsilon(r) = \kappa(r), \\
 & = & u Lip\, f + v Lip\, g + uv Lip\, f \cdot Lip\, g, & \textit{then denote} & \varepsilon(r) = \lambda(r), \\
 & = & u Lip\, f + v Lip\, g + uv Lip\, f \cdot Lip\, g + u + v, & \textit{then denote} & \varepsilon(r) = \omega(r).
\end{array}
$$

Noncausal coefficients (symmetric Ψ, and $\mathcal{F} = \mathcal{G}$) fit to non-causal processes. The ω coefficients are introduced here in order to cover strong mixing conditions and as well function of mixing coefficients as introduced in Billingsley (1969). A simple extension of the previous definitions concerns the case of random fields. For generality, consider $(X_t)_{y\in\mathbb{T}}$ with $(\mathbb{T}, d)$ a metric space, simple examples are $\mathbb{T} = \mathbb{Z}^d$ or $\mathbb{R}^d$.

Definition 3.2 [16] *A process $X = (X_t)_{t\in\mathbb{T}}$ with values in $\mathbb{R}^d$ is (ε, Ψ)-weakly dependent process if :*

$$
\varepsilon(r) = \sup_{\substack{u,v \geqslant 1 \\ (\mathbf{i},\mathbf{j}) \in I(u,v,r) \\ f\in\mathcal{F}_u, g\in\mathcal{G}_v}} \frac{1}{\Psi(f,g)} \left| Cov\Big(f(X_{i_1},\dots,X_{i_u}), g(X_{j_1},\dots,X_{j_v})\Big)\right| \to_{r\to\infty} 0,
$$

here $I(u,v,r)$ is the set of multiindices $(\mathbf{i},\mathbf{j}) = (i_1,\dots i_u, j_1,\dots,j_v) \in \mathbb{T}^{u+v}$ such that $d(\{i_1,\dots,i_u\},\{j_1,\dots,j_v\}) = \inf_{1\leqslant a\leqslant u, 1\leqslant b\leqslant v} d(i_a, j_b) \geqslant r$.

In this setting the coefficients η, κ, λ and ω are defined as above. For simplicity we shall not precise each coefficient but definitions are straightforward up to evident changes. In the sequel $r \geqslant 0$ will denote an arbitrary integer.

Heredity results. The following results are useful for various statistical applications.

Lemma 3.1 *For $k \geqslant 1$ let $Y_t = (X_{t-k},\dots,X_t)$ then $\theta_Y(r) \leq k\theta_X(r-k)$ if $r \geqslant k$.*

Proposition 3.1 [2] *Let $(X_n)_{n\in\mathbb{Z}}$ be a sequence of $\mathbb{R}^k$-valued random variables. Let $p > 1$. We assume that there exists some constant $C > 0$ such that $\max_{1\leq i\leq k} \|X_i\|_p \leq C$. Let h be a function from $\mathbb{R}^k$ to $\mathbb{R}$ such that $h(0) = 0$ and for $x, y \in \mathbb{R}^k$, there exist a in $[1, p[$ and $c > 0$ such that $|h(x) - h(y)| \leq c|x-y|(1 + |x|^{a-1} + |y|^{a-1})$.*
We define the sequence $(Y_n)_{n\in\mathbb{Z}}$ by $Y_n = h(X_n)$, then,
– if $(X_n)_{n\in\mathbb{Z}}$ is θ-weakly dependent, then $(Y_n)_{n\in\mathbb{Z}}$ too, $\theta_Y(r) = \mathcal{O}\left(\theta(r)^{\frac{p-a}{p-1}}\right)$;
– if $(X_n)_{n\in\mathbb{Z}}$ is η-weakly dependent, so is $(Y_n)_{n\in\mathbb{Z}}$ and $\eta_Y(r) = \mathcal{O}\left(\eta(r)^{\frac{p-a}{p-1}}\right)$;
– if $(X_n)_{n\in\mathbb{Z}}$ is λ-weakly dependent, $(Y_n)_{n\in\mathbb{Z}}$ also, $\lambda_Y(r) = \mathcal{O}\left(\lambda(r)^{\frac{p-a}{p+a-2}}\right)$.

More causal coefficients. For causal coefficients $\mathcal{F}$ will be the set of bounded functions and (usually) $\mathcal{G}$ will denote the set of Lipschitz functions.

– θ coefficients correspond to $\Psi(f,g) = v\|f\|_\infty \mathrm{Lip}\,(g)$, we define

$$\theta_p(\mathcal{M}, X) = \sup\{\|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}g(X)\|_p \,/\mathrm{Lip}\, g \leqslant 1\},$$

and then if $(X_i)_{i\in\mathbb{Z}}$ is an $\mathbb{L}^p$-sequence, and $(\mathcal{M}_k)_{k\in\mathbb{Z}}$ are σ-algebras $(\sigma(X_j, j \le k))$.

$$\theta_{p,v}(r) = \max_{s\leqslant v} \frac{1}{s} \sup_{i+r\le j_1 \le \cdots \le j_s} \theta_p\left(\mathcal{M}_i, (X_{j_1}, \ldots, X_{j_s})\right),\ \theta(r) = \theta_{1,\infty}(r)$$

– τ coefficients : $\tau_p(\mathcal{M}, X) = \left\| \sup\left\{ \int g(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int g(x)\mathbb{P}_X(dx) \,\big|\mathrm{Lip}\, g \leqslant 1\right\}\right\|_p$, and $\theta_p(\mathcal{M}, X) \le \tau_p(\mathcal{M}, X)$, now $\tau_{p,v}(r) = \max_{s\leqslant v} \frac{1}{s} \sup_{i+r\le j_1 \le \cdots \le j_s} \tau_p\left(\mathcal{M}_i, (X_{j_1}, \ldots, X_{j_v})\right).$

– γ-coefficients (projective measure) here, $\gamma_p(\mathcal{M}, X) = \|\mathbb{E}(X|\mathcal{M}) - \mathbb{E}(X)\|_p\,(\le\ \theta_p(\mathcal{M}, X))$ and, $\gamma_p(r) = \sup_{i\in\mathbb{Z}} \gamma_p(\mathcal{M}_i, X_{i+r})$.

3.2.4 Basic examples

A simple example writes $X_t = \sum_{j=-\infty}^{\infty} a_j \xi_{t-j}$ with $\mathbb{E}|\xi_0| \sum_{j=-\infty}^{\infty} |a_j| < \infty$ and $(\xi_i)_{i\in\mathbb{Z}}$ iid. More generally

$$X_t = H((\xi_{t-j})_{j\in\mathbb{Z}}), \qquad H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R},$$

where, more precisely, $H \in \mathbb{L}^m(\mu)$ for some $m \geqslant 0$, with μ the distribution of $(\xi_t)_{t\in\mathbb{Z}}$.

Proposition 3.2 [16] *The process $(X_t)_{t\in\mathbb{Z}}$ is η-weak dependent with $\eta(r) = 2\delta_{[r/2]}^{m\wedge 1}$ if*

$$\mathbb{E}\big|H\big(\xi_j, j\in\mathbb{Z}\big) - H\big(\xi_j 1\!\!1_{|j|<r}, j \in \mathbb{Z}\big)\big| \leqslant \delta_r \downarrow 0 \quad (r \uparrow \infty)$$

If $H\,(x_j, j\in\mathbb{Z})$ does not depend on x_j, $j < 0$, it is θ-dependent and $\theta(r) = \delta_r^{m\wedge 1}$.

Note. Even if this example looks general, it seems difficult to exhibit such functions H. We aim at describing more natural examples below.

Proposition 3.3 [16] *Gaussian or associated $\mathbb{L}^2-$processes are weakly dependent if*

$$\kappa(r) = \mathcal{O}\Big(\sup_{i\ge r} |Cov(X_0, X_i)|\Big) \to_{r\to\infty} 0.$$

3.2.5 Botanic of the models

We first mention LARCH(∞) models from Doukhan, Teyssière and Winant, [7] :

$$X_t = \xi_t\Big(a + \sum_{j=1}^{\infty} a_j X_{t-j}\Big) \tag{3.3}$$

here X_t is $d\times 1$, ξ_t is $d\times k$, a is $k\times 1$, and a_j are $k \times d$ matrices. Examples of such models follow.

– Bilinear (Giraitis, Surgailis, 2002, [24]) $X_t = \zeta_t\Big(a + \sum_{j=1}^{\infty} a_j X_{t-j}\Big) + b + \sum_{j=1}^{\infty} b_j X_{t-j}$

– GARCH(p,q) (Engle, 1982, [22]) $R_t = \sigma_t\varepsilon_t,\ \sigma_t^2 = \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 + \gamma_0 + \sum_{j=1}^{q}\gamma_j R_{t-j}^2$

– ARCH(∞) (Surgailis *et al.* 2000, [23])

$$R_t = \sigma_t\varepsilon_t, \qquad \sigma_t^2 = \beta_0 + \sum_{j=1}^{\infty}\beta_j R_{t-j}^2 \tag{3.4}$$

E.g. to write the model as a LARCH(∞) in this case, only set $\xi_t = \begin{pmatrix} \varepsilon_t & 1 \end{pmatrix}$, $a = \begin{pmatrix} \kappa\beta_0 \\ \lambda_1\beta_0 \end{pmatrix}$,
$a_j = \begin{pmatrix} \kappa\beta_j \\ \lambda_1\beta_j \end{pmatrix}$ with $\lambda_1 = \mathbb{E}(\varepsilon_0^2)$, $\kappa^2 = \mathbb{V}\mathrm{ar}(\varepsilon_0^2)$.
Note that $\phi = \|\xi_0\|_m \sum_j \|a_j\| < 1$, yields a solution of (3.3) :

$$X_t = \xi_t\Big(a + \sum_{k=1}^{\infty}\sum_{j_1,\ldots,j_k\geqslant 1} a_{j_1}\xi_{t-j_1}\cdots a_{j_k}\xi_{t-j_1-\cdots-j_k}a\Big) \tag{3.5}$$

If $Y_t \in \mathbb{L}^m$ is a solution of (3.3) for some $m \geqslant 1$, independent of ξ_s for $s > t$, then $Y_t = X_t$ *a.s.* (3.5). Moreover $\theta(r) \leqslant \|X_r - \tilde{X}_r\|_1$ and $\tau_{m,\infty}(t) \leqslant \|X_t - \tilde{X}_t\|_m$ with $A(s) = \sum_{j\geqslant s}\|a_j\|$,

$$\|X_t - \tilde{X}_t\|_m \leqslant \|\xi_0\|_m\Big(\|\xi_0\|_m \sum_{j<t} j\phi^{j-1}A\left(\frac{t}{j}\right) + \frac{\phi^t}{1-\phi}\Big) \tag{3.6}$$

e.g. $A(s) \leqslant Cs^{-b}$ or $\leqslant Cq^s$, imply that this expression $\leqslant C'/t^b$, $C'(q\vee\phi)^{\sqrt{t}}$.

Various generalizations of such models may also be found in [11], among them one may replace products by the Steutel & Van-Harn operator, $a \circ x = \mathrm{sign}(x)\sum_{i=1}^{x} Y_i$ for context-free and iid r.v Y_i with mean a ; which means that variables Y_i are choosen stochastically independent of the other random variables considered in the current problem and with $\mathbb{E}Y_i = a$. The previous models thus extend to a large class of integer valued random processes. Another extension is provided in the next section.

Models with infinite memory. Let $(\xi_t)_{t\in\mathbb{Z}}$ be iid, and $F : (\mathbb{R}^d)^{\mathbb{N}} \times \mathbb{R}^D \to \mathbb{R}^d$, we consider

$$X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \ldots ; \xi_t) \tag{3.7}$$

Theorem 3.1 [21] *Assume, for some $m \geqslant 1$, that $A = \|F(0,0,0,\ldots;\xi_t)\|_m < \infty$ and*
$$\|F(x_1,x_2,x_3,\ldots;\xi_t) - F(y_1,y_2,y_3,\ldots;\xi_t)\|_m \leqslant \sum_{j=1}^{\infty} a_j\|x_j - y_j\|.$$
Then existence of the model holds in $\mathbb{L}^m$, as well as its stationarity and its weak dependence with, $\theta(r) \leqslant C\inf_{N>0}\Big(\sum_{j\geqslant N} a_j + e^{-\alpha r/N}\Big)$ if $e^{-\alpha} = \sum_{j=1}^{\infty} a_j < 1$.

Random fields with infinite memory

$$X_t = F\Big((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_t\Big) \tag{3.8}$$

Existence is proved in [19] through Lipschitz type conditions as well as η type dependence with analogue decays as before.
A causal set is a set $C\subset\mathbb{Z}^d$ such that the convex cone generated by C does not contain the origin. Singletons, half lines or an open half space with half of its boundary in $\mathbb{Z}^2$ are causal, this improves on the usual quadrants condition.

- Regression models are defined by the equation
$$X_t = f(X_{t-\ell_1},\dots,X_{t-\ell_k}) + h(\xi_t)g(X_{t-m_1},\dots,X_{t-m_l}) + \xi_t.$$
If $\|f(x_1,\dots,x_k)-f(y_1,\dots,y_k)\| \leqslant \sum_{i=1}^k b_i\|x_i-y_i\|$ and $\|g(x_1,\dots,x_l)-g(y_1,\dots,y_l)\| \le \sum_{j=1}^l c_j\|x_j-y_j\|$ then $\eta(r)\leqslant C\big(e^{-\frac{\alpha}{2k}}\big)^r$ if, either $e^{-\alpha}=\sum_{i=1}^k b_i+\|h\|_\infty\sum_{i=1}^k c_i<1$, or $e^{-\alpha}=\sum_{i=1}^k b_i+\|h(\xi_0)\|_m\sum_{i=1}^k c_i<1$ if $\{\ell_1,\dots,\ell_k,m_1,\dots,m_l\}$ is a causal set.
- LARCH(∞) random fields satisfy $X_t=\xi_t\left(a+\sum_{j\in C}a_jX_{t-j}\right)$, $(t\in\mathbb{Z}^d)$; existence and weak dependence holds for ξ_t and a_j matrices, $\|\xi_0\|_\infty\sum_{j\in C}\|a_j\|<1$ for a norm of algebra; $\|\xi_0\|_m\sum_{j\in C}\|a_j\|<1$ is enough if C is causal.

Models with dependent innovations Set $X_t = H((\xi_{t-j})_{j\in\mathbb{Z}}$ for ξ a stationary process and we assume this process to be either η or λ-weak dependent.

Proposition 3.4 [20] *If $\mathbb{E}|\xi_0|^{m'}<\infty$, and assume that $x_j=y_j$ for $j\neq s$ implies $|H((x_j)_{j\in\mathbb{Z}})-H((y_j)_{j\in\mathbb{Z}})|\leqslant b_s\Big(\sup_{j\neq s}|x_j|^\ell\vee 1\Big)|x_s-y_s|$, then $\sum_s sb_s<\infty$, implies existence in $\mathbb{L}^m$, if $\ell m+1\leqslant m'$. If, moreover, $b_s\leqslant Cs^{-b}$ then, $\eta_\xi(r)\leqslant Cr^{-\eta}\Rightarrow \eta(r)\leqslant C'r^{-\eta\left(1-\frac{1}{b-1}\right)\frac{m'-2}{m'-\ell-1}}$ and $\lambda_\xi(r)\leqslant Cr^{-\lambda}\Rightarrow \lambda(r)\leqslant C'r^{-\lambda\left(1-\frac{2}{b}\right)\left(1-\frac{\ell}{m'-1}\right)}$.*

3.3 Least squares estimation of ARCH(∞) processes

In this section, we focus on the ARCH(∞) model (see § 3.2.5). We observe a process R_t solution of the equation 3.4. The least squares estimation method is studied and we use weak dependence to derive the asymptotic distribution of this estimate. Since we will only estimate a finite number of parameters, we fix the β_j for $j>j_0$: these are known parameters in the section and we shall assume that $\beta_k=\mathcal{O}(k^{-b})$ for some known $b>0$ then $\theta(t)\leqslant C't^{1-b}$ (see [7]). We will investigate how such condition can be weakened at the end of the section. The goal is to estimate a finite dimensional parameter $\beta=[\beta_0,\dots,\beta_{j_0}]$ by using a generated sample of size n.
Classical estimation procedures are least squares and maximum likelihood. The first one relies on the autoregressive representation of the square of the process but does not take heteroskedasticity into account. The second uses the form of the heteroskedasticity since it is the GLS. When R_t is $p-$Markovian (that is $(\beta_i)_{i\geqslant 0}\in\mathbb{R}^{(\mathbb{N})}$ admits only finitely many nonzero coefficients, leading to the

$ARCH(p)$ model), it is asymptotically efficient under strong hypothesis on the law chosen for the residuals. Hence it is easier to give general hypothesis for OLS under which its properties are known. The main advantage of OLS is its simplicity. For example when the number of parameters to be estimated increases, MLE becomes very hard to obtain in practice while OLS is quite as easy to compute. Another procedure is based on autocovariance of the square of the process, it is Whittle's estimate discussed in [2].

3.3.1 Definition, identifiability, existence and consistency

The square of an ARCH process has the AR(∞) representation $R_t^2 = \beta_0 + \sum_{j=1}^{\infty} \beta_j R_{t-j}^2 + \omega_t$, with $\omega_t = R_t^2 - \sigma_t^2$. Denote $z_t = \left(1, R_{t-1}^2, \ldots, R_{t-j_0}^2\right)'$ and $y_t = R_t^2 - \sum_{j=j_0+1}^{\infty} \beta_j R_{t-j}^2$. Since ω_t is the strong innovation of R_t^2, we have the identifiability conditions $\mathbb{E} z_t' \omega_t = 0 \Leftrightarrow \mathbb{E}(z_t' z_t)\beta = \mathbb{E}(z_t' y_t)$ and the regression model writes $y_t = z_t'\beta + \omega_t$. If Z_n and Y_n are the piled matrices in time, plugging the empirical law in the previous equation leads to the OLS estimator $\widehat{\beta}_{OLS} = (Z_n' Z_n)^{-1} Z_n' Y_n$.

Lemma 3.2 $\mathbb{P}(\varepsilon_t^2 = 1) \neq 1 \Rightarrow \mathbb{E} z_t' z_t$ *is positive definite, and hence so a.s. is* $\frac{1}{n}(Z_n' Z_n)$ *for large enough* n*, so that* $\widehat{\beta}_{OLS}$ *exists a.s.*

Proposition 3.2 *Under the hypothesis* $\mathbb{E} R_t^4 < \infty$, $\widehat{\beta}_{OLS} \to \beta$ *a.s. when* $n \to \infty$.

Proof. From the ergodic theorem and from heredity properties of ergodicity we obtain : $\frac{1}{n} Z_n' Z_n \to \mathbb{E} z_t' z_t$ and $\frac{1}{n} Z_n' Y_n \to \mathbb{E} z_t' y_t$.

3.3.2 Asymptotic normality

Asymptotic normality of the previous estimator requires both a CLT for the numerator $\frac{1}{\sqrt{n}} Z_n' Y_n$ and the *a.s.* convergence of denominator $n(Z_n' Z_n j)^{-1}$ to $\mathbb{E}(z_t' z_t)^{-1}$. When $\mathbb{E} R_t^8 < \infty$, the second condition is obtained via the ergodic theorem together with the continuity of the matrix inverse (since the lemma 3.2 shows that the limit is *a.s.* definite positive). The first condition is now proved by using weak dependence. We need to show asymptotic normality of the vector whose k-th component (for $k \in [0, j_0]$) is $(n^{-1/2} Z_n' Y_n)_{(K)} = n^{-1/2} \sum_{t=1}^{n} (R_t^2 - \sum_{j=j_0+1}^{\infty} \beta_j R_{t-j}^2)\,(\mathbb{1}_{k=0} + \mathbb{1}_{k \neq 0} R_{t-k}^2)$. We first recall Donsker type results, stated here simply for application sake, they also imply CLTs.

Theorem 3.1 (Donsker invariance principles, [10], [20]) *Let* $(X_i)_{i \in \mathbb{N}}$ *stationary 0-mean, with* $(2+\delta)$*-order moments* ($\delta > 0$)*. Assume that :*

- $(X_i)_{i \in \mathbb{N}}$ *is a* θ*-weakly dependenct times series with* $\theta(r) = \mathcal{O}(r^{-\theta})$, $\theta > 1 + 1/\delta$,
- $(X_i)_{i \in \mathbb{N}}$ *is* κ*-weakly dependent with* $\kappa(r) = \mathcal{O}(r^{-\kappa})$, $\kappa > 2 + 1/\delta$, *or*
- $(X_i)_{i \in \mathbb{N}}$ *is* λ*-weakly dependent,* $\lambda(r) = \mathcal{O}(r^{-\lambda})$, *with* $\lambda > 4 + 2/\delta$,

then $\displaystyle\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]} X_i \xrightarrow{\mathcal{L}}_{n\to\infty} \sigma W_t$ *in* $D(0,1)$*, Skorohod space, with* $\displaystyle\sigma^2 = \sum_{k=-\infty}^{\infty} \mathbb{E}X_0X_k \geqslant 0$.

Lemma 3.3 *Let* λ *be in* $\mathbb{R}^{j_0+1}$ *and* $u_t = \big(R_t^2 - \sum_{j=j_0+1}^{\infty}\beta_j R_{t-j}^2\big)\big(\lambda_0 + \sum_{k=1}^{j_0}\lambda_k R_{t-k}^2\big)$*. If* $\mathbb{E}R_t^{2p}$ *is finite then* u *is* θ*-weakly dependent with* $\theta_u(r) = \mathcal{O}\big(r^{-\frac{2p-2}{2p-1}\frac{p-2}{p-1}\frac{(b-1)^2}{b}}\big)$.

Proof. Let $w_t^{(M)} = \Big(R_t^2 - \sum_{j=j_0+1}^{M}\beta_j R_{t-j}^2\ ,\ \lambda_0 + \sum\limits_{k=1}^{j_0}\lambda_k R_{t-k}^2\Big)$, $w_t = w_t^{(\infty)}$, and for $t_1 \le \cdots \le t_u < s_1 \le \cdots \le s_v$ we set $w_{\mathbf{t}} = (w_{t_1},\dots,w_{t_u})$, $w_{\mathbf{s}} = (w_{s_1},\dots,w_{s_v})$, $w_{\mathbf{t}}^{(M)} = (w_{t_1}^{(M)},\dots,w_{t_u}^{(M)})$ and $w_{\mathbf{s}}^{(M)} = (w_{s_1}^{(M)},\dots,w_{s_v}^{(M)})$; we assume that $r = s_1 - t_u$, $f:\mathbb{R}^{2u}\to\mathbb{R}$ is bounded, and $g:\mathbb{R}^{2v}\to\mathbb{R}$ is Lipschitz then we write

$$\mathbb{C}\mathrm{ov}(f(w_{\mathbf{t}}),g(w_{\mathbf{s}})) = \mathbb{C}\mathrm{ov}(f(w_{\mathbf{t}}),g(w_{\mathbf{s}}) - g(w_{\mathbf{s}}^{(M)})) + \mathbb{C}\mathrm{ov}(f(w_{\mathbf{t}}),g(w_{\mathbf{s}}^{(M)})).$$

For the first term, we use

$$\begin{aligned}
\frac{|\mathbb{C}\mathrm{ov}(f(w_{\mathbf{t}}) - f(w_{\mathbf{t}}^{(M)}),g(w_{\mathbf{s}}))|}{v\mathrm{Lip}\,(g)\|f\|_\infty} &\le \frac{2}{v}\mathbb{E}\|w_{\mathbf{s}} - w_{\mathbf{s}}^{(M)}\| \le \frac{2}{v}\mathbb{E}\|\sum_{i=1}^{v} w_{s_i} - w_{s_i}^{(M)}\| \\
&\le \frac{2}{v}v\mathbb{E}|\sum_{j=M+1}^{\infty}\beta_j R_{s-j}^2| \le 2\mathbb{E}r_0^2\sum_{j=M+1}^{\infty}|\beta_j| \\
&= \mathcal{O}\big(A(M+1)\big) = \mathcal{O}(M^{-(b-1)})
\end{aligned}$$

For a process X, denote by θ_X any number a such that $\theta_X(r) = \mathcal{O}(r^{-a})$. We know that $\theta_R = b-1$. Applying Proposition 3.1 leads to $\theta_{R^2} = \frac{2p-2}{2p-1}\theta_R$. Because $g(w_s^{(M)})$ can be identified with a function $g(R_i^2,\dots,R_{i-M}^2; i\in s)$, we also have if $M \le r$,

$$\begin{aligned}
|\mathbb{C}\mathrm{ov}(f(w_{\mathbf{t}}),g(w_{\mathbf{s}}^{(M)}))| &\le Mv\mathrm{Lip}\,(g)||f||_\infty\theta_{R^2}(r-M) \\
&= \mathcal{O}(M(r-M)^{-\theta_{R^2}})
\end{aligned}$$

In order to have the same rate in both terms, we set $M = r^{\theta_{R^2}/b} = r^{\frac{2p-2}{2p-1}\frac{b-1}{b}}$. Then $M = o(r)$ and the resulting rate is $\theta_w = \frac{2p-2}{2p-1}\frac{(b-1)^2}{b}$. Applying again Proposition 3.1 to the product of the component of w, we find that $\theta_u = \frac{p-2}{p-1}\,\theta_w = \frac{p-2}{p-1}\frac{2p-2}{2p-1}\frac{(b-1)^2}{b}$. □

Finally, we find the moment condition required to match with the theorem 3.1. In the next theorem, the asymptotic normality of $\widehat{\beta}_{OLS}$ is obtained only for b above a constant level. The required moment on R tends to 0 as b grows, that is as the coefficients β_j tends more quickly to 0.

Theorem 3.2 *Assume that* $b > \frac{3+\sqrt{5}}{2}$*. Let* $\delta > \frac{7}{4}\frac{b}{b^2-3b+1} > 0$*. If* $\mathbb{E}R_t^{8+4\delta} < \infty$*, then* $\sqrt{n}(\widehat{\beta}_{OLS} - \beta)$ *is asymptotically Gaussian with mean 0.*

Proof. Let $\phi(b) = \frac{(b-1)^2}{b} = b - 2 + \frac{1}{b}$ which grows from 0 to ∞ with $b > 1$. We apply the previous lemma with $p = 4 + 2\delta$. This leads to $\theta_u = \phi(b)\frac{4(\delta+1)}{4\delta+7}$. To apply Theorem 3.1 we check : $\theta_u > (\delta+1)/\delta \Leftrightarrow \phi(b) > 1 + \dfrac{7}{4\delta} \Leftrightarrow \delta > \dfrac{7}{4}(\phi(b)-1)^{-1}$ and $\phi(b) > 1$.
Finally $\dfrac{1}{\sqrt{n}}\displaystyle\sum_{t=1}^{n} \lambda'(z_t' z_t) = \dfrac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t$ is asymptotically Gaussian ($\forall \lambda \in \mathbb{R}^{j_0+1}$). □

The last issue is the calculability. The estimator $\widehat{\beta}_{OLS}$ depends on the infinite past of the process which is never observed in practice. The real world estimator has the same expression but replacing y_t by $\widetilde{y}_t = R_t^2 - \sum_{j=j_0+1}^{t} \beta_j R_{t-j}^2$, that is truncating the sum to the observed data. Finally $\widetilde{\beta}_{OLS} = (Z_n' Z_n)^{-1} Z_n' \tilde{Y}_n$ with the matrices $(Z)_{(t,k)} = \mathbb{1}_{k\neq 0} R_{t-k}^2 + \mathbb{1}_{k=0}$ and $(\tilde{Y})_{(t,1)} = \tilde{y}_t$. The next theorem shows that the difference is negligible.

Theorem 3.3 *Under the assumptions of Theorem 3.2, $\sqrt{n}(\widetilde{\beta}_{OLS} - \beta)$ is asymptotically Gaussian with mean 0.*

Proof. We show that $n^{-1/2} Z_n' \widetilde{Y}_n$ and $n^{-1/2} Z_n' Y_n$ have the same limit in law.
Let $\Delta_t = z_t'(y_t - \widetilde{y}_t)$, we have

$$\begin{aligned} \mathbb{E}|\Delta_t| &\leq |\mathbb{C}\mathrm{ov}(Y_t - \widetilde{Y}_t, Z_t)| + \mathbb{E}|Y_t - \widetilde{Y}_t|\,\mathbb{E}|Z_t| \\ &= \mathcal{O}(\theta_{r^2}(t) + A(t+1)) \\ &= \mathcal{O}(t^{-(b-1)\frac{2p-2}{2p-1}}) \end{aligned}$$

It implies $\mathbb{E}|n^{-1/2}\sum_{t=1}^{n} \Delta_t| = \mathcal{O}(n^{-1/2}(n-1)^{-(b-1)\frac{2p-2}{2p-1}+1})$ so that

$$p > 1 + \frac{1}{2(b-3)} \Rightarrow \frac{2p-2}{2p-1}(b-1) > 1/2 \Rightarrow \mathbb{E}\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \Delta_t\Big| = o(1). \ \square$$

Another result of importance for statistical applications is the following empirical CLT.

Theorem 3.4 (Empirical CLT [11]) *Assume that X_n is stationary with uniform marginal distributions, then $\dfrac{1}{\sqrt{n}}\displaystyle\sum_{k=1}^{n}(\mathbf{1}(X_k \leqslant x) - F(x)) \xrightarrow{D[\mathbb{R}]}_{n\to\infty} Z(x)$ where the centered Gaussian process $(Z(x))_{x\in\mathbb{R}}$ admits the covariance*

$$\Gamma(x,y) = \sum_{k=-\infty}^{\infty} \mathbb{C}\mathrm{ov}(\,\mathbf{1}(X_0 \leqslant x), \mathbf{1}(X_k \leqslant y)),$$

if one of the following weak dependence condition holds,
- $\theta(i) = \mathcal{O}(i^{-a})$ *for* $a > 2 + 2\sqrt{2} \approx 4.8\cdots$,
- $\eta(i) = \mathcal{O}(i^{-a})$ *for* $a > 5$,
- $\lambda(i) = \mathcal{O}(i^{-a})$ *for* $a > 15/2$.

Note that coupling arguments also yield ECLT for classes of functions with polynomial bracketing entropy decay conditions, see [12] and [11]. Besides evident applications to empirical statistical techniques and to functional estimation questions (also discussed in [11]), this yields a tool to derive asymptotics of MLE estimates.

3.3.3 Effective estimation procedures

Knowing the parameters with index larger than j_0 is of course an unrealistic hypothesis excepted for $p-$Markov ARCH(p) processes. In the general setting, if the β are unknown, we cannot remove the exact impact of their lags in y_t. We can only regress R_t^2 on a truncated past of j_0 lags, with $x_t = R_t^2$ instead of y_t. This lack of explanatory variables biases the OLS estimator as long as j_0 remains fixed. Using the same counting as above, the error is $|\Delta_t| = \mathcal{O}(j_0^{-(b-1)\frac{2p-2}{2p-1}})$. So we will need to set $j_0 = j_0(n) \to \infty$. The real problem is not non-parametricity but the fact that matrices dimensions should increase. For example in a parametric setting $\beta = \phi(\theta)$, the matrices still have growing dimension. For this asymptotics, not much is given by the classical OLS theory.
Hence it does not seem harder to estimate β than just its first components. This estimation problem is straightaway a nonparametric one. It is much like the density estimation problem : when one wants to estimate $f(x_0)$ with kernels, the variance has a non parametric $\sqrt{n/j_0(n)}$ while bias is $\mathcal{O}(j_0^{-\rho}(n))$ under ρ-regularity conditions if we set $1/j_0(n) \to 0$ for the corresponding window width. The idea is thus to achieve analogues nonparametric rates for our OLS procedures, but how can we derive rates of convergence for matrices with growing size ?
First we may use weak dependence to obtain $\mathbb{L}^2$ rates of convergence for coordinates of the numerator by using the following ideas. We consider $j > 0$ because of notations, but the case $j = 0$ is in fact easier ; for the component j, $A = \frac{1}{n}(Z'X)_{(j)}$, we have, $\mathbb{V}\mathrm{ar}(A) \le \frac{1}{n^2}\sum_{t,s\in[1,n]} |\mathbb{C}\mathrm{ov}(x_t x_{t-j}, x_s x_{s-j})|$, and we use similar notations for $X_n = [x_1, \dots, x_n]$. We divide this sum respectively for indices belonging to Δ and to Δ^c with $\Delta = \{i, j/|i-j| \le t_0\}$. On Δ, each covariance is bounded by a moment :

$$|\mathbb{C}\mathrm{ov}(x_t x_{t-j}, x_s x_{s-j})| \le \mathrm{Var}\,(x_t x_{t-j}) \le \mathrm{Var}\, x_t^2 + (\mathbb{E}x_t^2)^2 + (\mathbb{E}x_t^2 + \mathrm{Var}\, x_t)^2,$$

while on Δ^c we will sum bounds proved from weak dependence :

$$|\mathbb{C}\mathrm{ov}(x_t x_{t-j}, x_s x_{s-j})| \le \theta_{x_t x_{t-j}}(|t-s|) \le \theta_{[x_t,\dots,x_{t-j}]}(|t-s|)^{\frac{p-2}{p-1}} \le \theta_{x_t}(|t-s|-j)^{\frac{p-2}{p-1}}.$$

We thus obtain

$$\begin{aligned} \mathbb{V}\mathrm{ar}(A) &\le \frac{\#\Delta}{n^2} + \frac{1}{n^2}\sum_{t,s\in\Delta^c} \theta_x(|t-s|-j)^{\frac{p-2}{p-1}} \\ &= \mathcal{O}\Big(\frac{t_0+1}{n} + \frac{1}{n^2}\sum_{t=t_0+1}^{T}(T-t)\theta_x(t-j)^{\frac{p-2}{p-1}}\Big) \end{aligned}$$

Then, we need to sum over j to derive the rate of convergence for the whole numerator. We can only sum over the first values of j to obtain rate for the first components of β.
The same approach can be used for the denominator, but because we need to invert and multiply it, we need to link the rate obtained in the $\mathbb{L}^1$ norm to a multiplicative one ($e.g.$ $|||M|||_1 \equiv \sum_i \max_j |M_{i,j}| \leq \|M\|_1 \equiv \sum_i \sum_j |M_{i,j}|$). For the inversion we use

$$(Z_n' Z_n)^{-1} - (\mathbb{E}z_t' z_t)^{-1} = -(Z_n' Z_n)^{-1}(Z_n' Z_n - \mathbb{E}z_t' z_t)\mathbb{E}(z_t' z_t)^{-1},$$

however there is still a lot of technical points to adress here. The balance of t_0 and j_0 as functions of n leading to the optimal rate is reported to a future paper. Analogously to kernel density estimations a marginal CLT with normalization $\sqrt{n/j_0(n)}$ will be exhibited.

3.4 Random fields

We now consider a real-valued, centered and η, λ or $\omega-$weakly dependent random field X such that $M = \sup_{t\in\mathbb{Z}^d} \mathbb{E}|X_t|^m < \infty$ for a real number $m > 2$. We use here weak dependence in order to derive moments inequalities and a central limit theorem as well as a Donsker type invariance principle for partial sums of a stationary random field.
The following result provides covariances inequalities used in the proof of forthcoming moments inequalities and of a central limit theorem.

Lemma 3.1 *Let X be a real valued and centered $\eta-$weakly dependent random field. If $M = \sup_{t\in\mathbb{Z}^d} \mathbb{E}|X_t|^m < \infty$ for some real number $m > 2$, then*

1. *If X is $\lambda-$weakly dependent : $|Cov(X_i, X_j)| \leq 10 M^{\frac{1}{m(m-1)}} \lambda_X(|i-j|)^{\frac{m-2}{m-1}}$.*
   Moreover $\sup_{i\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d} |Cov(X_i, X_j)| < \infty$ if $\lambda_X(r) = \mathcal{O}\left(r^{-\lambda}\right)$ with $\lambda > d\frac{m-1}{m-2}$.
2. *If X is $\omega-$weakly dependent : $|Cov(X_i, X_j)| \leq 14 M^{\frac{2}{m^2}} \omega_X(|i-j|)^{\frac{m-2}{m}}$.*
   Moreover $\sup_{i\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d} |Cov(X_i, X_j)| < \infty$ if $\omega_X(r) = \mathcal{O}\left(r^{-\omega}\right)$ with $\omega > d\frac{m}{m-2}$.
3. *Suppose $\sup_{t\in\mathbb{Z}^d} \|X_t\|_\infty < \infty$ and X is $\lambda-$weakly dependent (resp. $\omega-$weakly dependent). If $\lambda_X(r) = \mathcal{O}\left(r^{-\lambda}\right)$ with $\lambda > d$ (resp. $\omega_X(r) = \mathcal{O}\left(r^{-\omega}\right)$ with $\omega > d$) then $\sup_{i\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d} |Cov(X_i, X_j)| < \infty$.*

Proof. We use a truncation technique. Let $T \geq 1$. We denote for $i \in \mathbb{Z}^d$, $X_{T,i} = (-T) \vee X_i \wedge T$ and

$$\overline{X}_i = X_{T,i} - \mathbb{E}X_{T,i}$$

Let $a \in [2, m]$, as $|X_i - X_{T,i}|^a \le |X_i|^a \, \mathbb{1}_{|X_i|>T}$, we obtain $\left\|X_i - \overline{X_i}\right\|_a \le 2M^{\frac{1}{a}} T^{1-\frac{m}{a}}$. Write $|\mathrm{Cov}(X_i, X_j)| \le \left|\mathrm{Cov}(X_i - \overline{X}_i, X_j)\right| + \left|\mathrm{Cov}(X_j - \overline{X}_j, \overline{X}_i)\right| + \left|\mathrm{Cov}(\overline{X}_i, \overline{X}_j)\right|$. If $1/a + 1/m = 1$, then

$$\begin{aligned}\left|\mathrm{Cov}(X_i - \overline{X}_i, X_j)\right| &+ \left|\mathrm{Cov}(X_j - \overline{X}_j, \overline{X}_i)\right| \\ &\le \left\|X_i - \overline{X}_i\right\|_a \|X_j\|_m + \|X_{T,i}\|_m \left\|X_j - \overline{X}_j\right\|_a \\ &\le 4MT^{2-m}\end{aligned}$$

1. Under $\lambda-$dependence, $|\mathrm{Cov}(\overline{X}_i, \overline{X_j})| \le 3T\lambda_X(|i-j|)$. As we may always assume that $\lambda(r) \leqslant 2M^{1/m}$, we get the result by setting $T = 2^{\frac{1}{m-1}} M^{\frac{1}{m(m-1)}} \lambda_X(|i-j|)^{-\frac{1}{m-1}}$.
2. For $\omega-$dependence, $|\mathrm{Cov}(X_i, X_j)| \le 4MT^{2-m} + 5T^2\omega_X(|i-j|)$. The result follows now with $T = 2^{\frac{1}{m}} M^{\frac{1}{m^2}} \omega_X(|i-j|)^{\frac{-1}{m}}$. The last point 3. is obvious. □

3.4.1 Moment inequalities for partial sums

This section is aimed at deriving moment inequalities of the Marcinkiewicz-Zygmund type, $\mathbb{E}\Big|\sum_{j\in U} X_j\Big|^q \le C\,|U|^{q/2}$, where U is a block set of $\mathbb{Z}^d$ (i.e $U = (a, b] = ((a_1, b_1] \times \cdots \times (a_d, b_d]) \cap \mathbb{Z}^d$ with $a, b \in \mathbb{R}^d, a_1 < b_1, \ldots, a_d < b_d$), $|U|$ stands for the cardinality of a finite set U and q is a real number ≥ 2. The proof adapts ideas in [9] for moments inequality with order $q = 2 + \delta$, $\delta \in [0, 1]$. For a block U, let $S(U) = \sum_{j\in U} X_j$. For a real number $s \ge 2$, we will say that X verifes condition $\mathbf{C}(s)$ if
$\mathbf{C}(s)$: $\exists C_s > 0$ such that for each block $U \subset \mathbb{Z}^d$, $\|S(U)\|_s \le C_s\,|U|^{1/2}$.
In this section, we suppose that X is either $\lambda-$weakly dependent or $\omega-$weakly dependent with

$$\lambda_X(r) \le K(r+1)^{-\lambda}, \qquad \lambda > d\frac{m-1}{m-2} \tag{3.9}$$

$$\omega_X(r) \le K(r+1)^{-\omega}, \qquad \omega > d\left(2 \vee \frac{m}{m-2}\right) \tag{3.10}$$

Now (3.9) (resp. (3.10)) asserts $\mathbf{C(2)}$ with $C_2 = \sup_{j\in\mathbb{Z}^d} \sum_{i\in\mathbb{Z}^d} |\mathrm{Cov}(X_i, X_j)|$ with Lemma 3.1.

The general case

We first prove a technical lemma in order to derive $\mathbf{C(q)} \Rightarrow \mathbf{C(q+\delta)}$.

Lemma 3.2 *Let $q \in \mathbb{N}$, $q \ge 2$, $\delta \in (0, 1]$ such that $\Delta = q + \delta < m$. We suppose that X verifies condition $\mathbf{C(q)}$.*
Under $\lambda-$dependence, if $\lambda \ge \frac{d}{2} \cdot \frac{m(q+2-\delta)+(\delta-2)\Delta}{m-\Delta}$, condition $\mathbf{C}(\boldsymbol{\Delta})$ holds.
In the case of $\omega-$dependence, $\omega \ge \frac{d}{2}\frac{m(q+6-\delta)+(\delta-5)\Delta}{m-\Delta}$, implies the condition $\mathbf{C}(\boldsymbol{\Delta})$.

Proof. We follow the proof of [9]. Let $h(n) = \min\{k \in \mathbb{N}/2^k \geq n\}$ and for a block $U \subset \mathbb{Z}^d$ having edges with lengths $l_1, \ldots, l_d$, we set $h(U) = h(l_1) + \ldots h(l_d)$. We prove the result by induction on $h(U)$. The result is obvious if $h(U) = 0 \Leftrightarrow |U| = 1$. We suppose the result true for a block U such that $h(U) \leq h_0$ and let U a block such that $h(U) = h_0 + 1$. Let L be any of the widest edges of U. Denote its lenght by $l(U)$. As pointed in Lemma 1 in [9], if we draw two hyperplanes orthogonal to L dividing U into two blocks U_1 and U_2 of thickness $[l(U)/2]$ and $l(U) - [l(U)/2]$ respectively, then there exists a value $\tau = \tau(\Delta) < 1$ independent of U such that :

$$|U_1|^{\Delta/2} + |U_2|^{\Delta/2} \leq \tau \, |U|^{\Delta/2}$$

(here $[\cdot]$ stands for the integer part of a number). Let $\xi \in (0, 1/2]$ such that $1 - \tau^{1/\Delta} - \sqrt{\xi} > 0$. We divide U into blocks in the following way :
i) If $\xi \, |U|^{1/d} \geq 1$, we draw two hyperplanes orthogonal to L, dividing U into three blocks, a central block V admits the thickness $[\xi \, |U|^{1/d}]$ and the two other blocks Q and R have thickness $\leq l(U)/2$. In some case Q or R may be empty.
ii) If $\xi \, |U|^{1/d} < 1$, we draw one hyperplane dividing U into two blocks Q and R of thickness $[l(U)/2]$ and $l(U) - [l(U)/2]$ respectively. We set $V = \emptyset$.
We have always :

$$|Q|^{\Delta/2} + |R|^{\Delta/2} \leq \tau \, |U|^{\Delta/2} . \tag{3.11}$$

Moreover $h(Q), h(R), h(V) \leq h_0$ and we may use induction. We set $r = d(R, Q)$. Then

$$\|S(Q) + S(V) + S(R)\|_\Delta \leq \|S(Q) + S(R)\|_\Delta + \|S(V)\|_\Delta$$

>From induction, $\|S(V)\|_\Delta \leq C_\Delta \, |V|^{1/2} \leq C_\Delta \sqrt{\xi} \, |U|^{1/2}$ for some constant $C_\Delta > 0$. Using Lemma 5.17 yields :

$$\begin{aligned} \mathbb{E} \, |S(Q) + S(R)|^\Delta \leq \; & \mathbb{E} \, |S(Q)|^\Delta + \mathbb{E} \, |S(R)|^\Delta \\ & + (2^{q+1} - 3) \left(\mathbb{E} \, |S(Q)|^q \, |S(R)|^\delta + \mathbb{E} \, |S(R)|^q \, |S(Q)|^\delta \right) \end{aligned}$$

1) We now use the induction for the first two terms :

$$\mathbb{E} \, |S(Q)|^\Delta + \mathbb{E} \, |S(R)|^\Delta \leq C_\Delta^\Delta \left(|Q|^{\Delta/2} + |R|^{\Delta/2} \right) \leq C_\Delta^\Delta \tau \, |U|^{\Delta/2}$$

2) For the third term set $\overline{S}(R) = (-T) \vee S(R) \wedge T$ for a truncation $T \geq |U|$ and use the inequality :

$$\begin{aligned} \mathbb{E} \, |S(Q)|^q \, |S(R)|^\delta & \leq \mathbb{E} \, |S(Q)|^q \left|\overline{S}(R)\right|^\delta + \mathbb{E} \, |S(Q)|^q \left|S(R) - \overline{S}(R)\right|^\delta \\ & \leq 2^{q-1} \mathbb{E} \left|S(Q) - \overline{S}(Q)\right|^q \left|\overline{S}(R)\right|^\delta \\ & \quad + 2^q \mathbb{E} \left|\overline{S}(Q)\right|^q \left|\overline{S}(R)\right|^\delta \\ & \quad + \mathbb{E} \, |S(Q)|^q \left|S(R) - \overline{S}(R)\right|^\delta \end{aligned}$$

– We first obtain

$$\begin{aligned}\mathbb{E}\left|S(Q)-\overline{S}(Q)\right|^q\left|\overline{S}(R)\right|^\delta &\le T^\delta \mathbb{E}\left|S(Q)\right|^q \mathbb{1}_{\{|S(Q)|\ge T\}}\\ &\le T^\delta T^{q-m}\mathbb{E}\left|S(Q)\right|^m\\ &\le MT^{\Delta-m}\left|U\right|^m\end{aligned}$$

– For the second term write a covariance :

$$\begin{aligned}\mathbb{E}\left|\overline{S}(Q)\right|^q\left|\overline{S}(R)\right|^\delta &\le \mathbb{E}\left|\overline{S}(Q)\right|^q\left((1+|S(R)|)^\delta \wedge T^\delta\right)\\ &= \mathbb{C}\mathrm{ov}\left(\left|\overline{S}(Q)\right|^q,\left((1+|S(R)|)^\delta\wedge T^\delta\right)\right)\\ &\quad +\mathbb{E}\left|\overline{S}(Q)\right|^q\mathbb{E}\left((1+|S(R)|)^\delta\wedge T^\delta\right)\end{aligned}$$

Observe that $f:(x_1,\dots,x_{|Q|})\mapsto \left|(-T)\vee\sum_{i=1}^{|Q|}x_i\wedge T\right|^q$ is $qT^{q-1}-$Lipschitz and bounded by T^q, and $g:(x_1,\dots,x_{|R|})\mapsto\left(1+\left|\sum_{i=1}^{|R|}x_i\right|\right)^\delta\wedge T^\delta$ is $\delta-$Lipschitz and it is bounded by T^δ. Under λ dependence this implies, if $T\ge|U|$,

$$\begin{aligned}\left|\mathrm{Cov}\left(\left|\bar{S}(Q)\right|^q,\left(1+\left|\bar{S}(R)\right|\right)^\delta\right)\right| &\le (qT^{q+\delta-1}\left|Q\right|+\delta T^q\left|R\right|+q\delta T^{q-1}\left|U\right|^2)\lambda_X(r)\\ &\le (2q+1)T^q\left|U\right|\lambda_X(r)\end{aligned}$$

Under $\omega-$dependence, if $T\ge|U|$, then

$$\begin{aligned}\left|\mathrm{Cov}\left(\left|\bar{S}(Q)\right|^q,\left(1+\left|\bar{S}(R)\right|\right)^\delta\right)\right| &\le \Big\{(qT^{q+\delta-1}+T^{q+\delta})|Q|\\ &\quad+ (\delta T^q+T^{q+\delta})|R|+q\delta T^{q-1}|U|^2\Big\}\omega_X(r)\\ &\le (2q+3)T^{q+1}\left|U\right|\omega_X(r)\end{aligned}$$

Moreover

$$\begin{aligned}\mathbb{E}\left|\bar{S}(Q)\right|^q\mathbb{E}\left(1+\left|\bar{S}(R)\right|\right)^\delta &\le C_q^q\left|Q\right|^{q/2}(1+\mathbb{E}\left|S(R)\right|^\delta\\ &\le C_q^q\left|Q\right|^{q/2}(1+\|S(R)\|_2^\delta)\end{aligned}$$

As $\|S(R)\|_2\le C_2\left|U\right|^{1/2}$, we have

$$\mathbb{E}\left|\overline{S}(Q)\right|^q\mathbb{E}\left(1+\left|\overline{S}(R)\right|\right)^\delta\le 2(C_2\vee 1)C_q^q\left|U\right|^{\Delta/2}$$

– For the last term, with $q/m+1/m^{'}=1$, we get :

$$\begin{aligned}\mathbb{E}\left|S(Q)\right|^q\left|S(R)-\overline{S}(R)\right|^\delta &\le \mathbb{E}\left|S(Q)\right|^q\left|S(R)\right|^\delta\mathbb{1}_{\{|S(R)|\ge T\}}\\ &\le \|S(Q)\|_m^q\left\|\left|S(R)\right|^\delta\mathbb{1}_{\{|S(R)|\ge T\}}\right\|_{m'}\\ &\le \left|Q\right|^q M^{q/m}T^{\delta-\frac{m}{m'}}\left\|S(R)\right\|_m^{m/m'}\\ &\le M\left|U\right|^{q+\frac{m}{m'}}T^{\delta-\frac{m}{m'}}\end{aligned}$$

As $m/m' = m - q$:

$$\mathbb{E}\,|S(Q)|^q \left|S(R) - \overline{S}(R)\right|^\delta \le M\,|U|^m\,T^{\Delta - m}$$

– Under $\lambda-$dependence we have proved

$$\begin{aligned}\mathbb{E}\,|S(Q)|^q\,|S(R)|^\delta &\le (2^q+1)M\,|U|^m\,T^{\Delta-m} + 2^q(2q+1)T^q\,|U|\,\lambda_X(r)\\ &+ 2^{q+1}(C_2\vee 1)C_q^q\,|U|^{\Delta/2}\end{aligned}$$

Observe that from (3.9) and from the definition of r : $\lambda_X(r) \le K\xi^{-\lambda}\,|U|^{-\lambda/d}$. Choose $T = |U|^{(m-1+\frac{\lambda}{d})/(m-\delta)}$. This leads to

$$\begin{aligned}\mathbb{E}\,|S(Q)|^q\,|S(R)|^\delta &\le \left[(2^q+1)M + 2^q(2q+1)K\xi^{-\eta}\right]|U|^{\frac{m(q+1)-\Delta}{m-\delta} - \frac{\lambda}{d}\frac{m-\Delta}{m-\delta}}\\ &+ 2^{q+1}(C_2\vee 1)C_q^q\,|U|^{\Delta/2}\end{aligned}$$

By assumption $\lambda \ge \frac{d}{2}\times\frac{m(q+2-\delta)+(\delta-2)\Delta}{m-\Delta}$ and so :

$$\mathbb{E}\,|S(Q)|^q\,|S(R)|^\delta \le \left[(2^q+1)M + 2^q(2q+1)K\xi^{-\eta} + 2^{q+1}(C_2\vee 1)C_q^q\right]|U|^{\Delta/2}$$

– Under $\omega-$dependence,

$$\begin{aligned}\mathbb{E}\,|S(Q)|^q\,|S(R)|^\delta &\le (2^q+1)M\,|U|^m\,T^{\Delta-m}\\ &+2^q(2q+3)T^{q+1}\,|U|\,\omega_X(r) + 2^{q+1}(C_2\vee 1)C_q^q\,|U|^{\Delta/2}\end{aligned}$$

Here from (3.10) and from the definition of $r : \omega_X(r) \le K\xi^{-\omega}\,|U|^{-\omega/d}$. Choose $T = |U|^{(m-1+\frac{\omega}{d}/(m+1-\delta)}$. Note that the assumption $\omega > 2d$ implies $\frac{m-1+\frac{\omega}{d}}{m+1-\delta} \ge 1$. This yields

$$\begin{aligned}\mathbb{E}\,|S(Q)|^q\,|S(R)|^\delta &\le \left[(2^q+1)M + 2^q(2q+3)K\xi^{-\eta}\right]|U|^{\frac{m(q+2)-\Delta}{m+1-\delta} - \frac{\omega}{d}\frac{m-\Delta}{m+1-\delta}}\\ &+2^{q+1}(C_2\vee 1)C_q^q\,|U|^{\Delta/2}\end{aligned}$$

By assumption $\omega \ge \frac{d}{2}\times\frac{m(q+4-\delta)+(\delta-3)\Delta}{m-\Delta}$, thus :

$$\mathbb{E}\,|S(Q)|^q\,|S(R)|^\delta \le \left[(2^q+1)M + 2^q(2q+3)K\xi^{-\eta} + 2^{q+1}(C_2\vee 1)C_q^q\right]|U|^{\Delta/2}$$

3) Using the same ideas for the last term, from inequality (3.11) we derive

$$\|S(U)\|_\Delta \le \left[C_\Delta\left(\sqrt{\xi} + \tau^{1/\Delta}\right) + L^{1/\Delta}\right]|U|^{1/2}$$

with respectively, in the $\lambda-$dependent case and the $\omega-$dependent case :

$$\begin{aligned}L &= 2(2^{q+1}-3)\left[(2^q+1)M + 2^q(2q+1)K\xi^{-\lambda} + 2^{q+1}(C_2\vee 1)C_q^q\right],\\ L &= 2(2^{q+1}-3)\left[(2^q+1)M + 2^q(2q+3)K\xi^{-\omega} + 2^{q+1}(C_2\vee 1)C_q^q\right]\end{aligned}.$$

The choice $C_\Delta \ge L^{1/\Delta}/(1-\tau^{1/\Delta}-\sqrt{\xi})$ yields the result. □

Corollary 3.1 *1. Under $\lambda-$weak dependence, $\mathbf{C(q)}$ holds with $q = \left[\frac{2\lambda m}{2\lambda+(m-1)d}\right]$. Suppose, moreover that $\lambda \ge \frac{d}{2}(m^2-m+2)$ if $m \in \mathbb{N}$ or $\lambda \ge \frac{d}{2}\frac{[m](m-2)+2m}{m-[m]}$ if $m \notin \mathbb{N}$, then condition $\mathbf{C(q+\delta)}$ holds with $0 \le \delta < 1 \wedge (m-q) \wedge \frac{-a+\sqrt{a^2-4bd}}{2d}$ where $a = 2\lambda - d(m+2-q)$ and $b = (m-2)qd + 2dm - 2(m-q)\lambda$.*

2. If X is $\omega-$weakly dependent, condition $\mathbf{C(q)}$ *holds with* $q = 2 \vee \left[\frac{2m(\omega-d)}{2\omega+(m-2)d}\right]$.
If moreover $\omega \geq \frac{d}{2}(m^2+3)$ *if* $m \in \mathbb{N}$ *or* $\omega \geq \frac{d}{2}\frac{[m](m-3)+4m}{m-[m]}$ *if* $m \notin \mathbb{N}$, *then condition* $\mathbf{C}(\mathbf{q}+\delta)$ *holds for* $0 \leq \delta < 1 \wedge (m-q) \wedge \frac{-a+\sqrt{a^2-4bd}}{2d}$ *where* $a = 2\omega - d(m+3-q)$ *and* $b = (m-3)qd + 4dm - 2(m-q)\omega$.

Proof. **1.** As condition $\mathbf{C}(2)$ holds, the result follows by induction on $\{2, \ldots, q\}$ by using Lemma 3.2. From Lemma 3.2, condition $\mathbf{C}(q+\delta)$ holds for $\lambda \geq \frac{d}{2}\frac{m(q+2-\delta)+(\delta-2)\Delta}{m-\Delta}$. This relation is rewritten $d\delta^2 + a\delta + b \leq 0$. If we prove that $b \leq 0$, the result will follow (because $\frac{-a+\sqrt{a^2-4bd}}{2d} \geq 0$). But $b \leq 0 \Leftrightarrow q \leq 2m\frac{\lambda-d}{2\eta+(m-2)d}$. If $m \in \mathbb{N}$, we have $q \leq m-1$ and $m-1 \leq 2m\frac{\lambda-d}{2\lambda+(m-2)d} \Leftrightarrow \lambda \geq \frac{d}{2}(m^2-m+2)$. If now $m \notin \mathbb{N}$, then $q \leq [m]$ and $[m] \leq 2m\frac{\lambda-d}{2\lambda+(m-2)d} \Leftrightarrow \frac{d}{2}\frac{[m](m-2)+2m}{m-[m]}$ which proves the result. The proof of **2.** is omitted. □

The case of bounded random variables

Lemma 3.3 *We suppose* $N = \sup_{t\in\mathbb{Z}^d} \|X_t\|_\infty < \infty$ *and* λ *(resp.* $\omega > d$*). Then condition* $\mathbf{C}(s)$ *holds with* $s = [2\lambda/d]$ *(resp.* $s = 2 \vee [2\omega/d - 2]$*).*

Proof. The result is obvious for $s = 2$ since $\lambda > d$ (resp. $\omega > d$) implies the relation $\sup_{i\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d} |\mathrm{Cov}(X_i,$ ∞. Assume the result if $s = q \geq 2$ where $q \in \mathbb{N}$, $q \leq s-1$ and $2\omega/3 - 2 \geq 3$. We derive condition $\mathbf{C}(q+1)$ as in the proof of Lemma 3.2. With the same notations, this is sufficient to modify only the step 2) :

$$\mathbb{E}\,|S(Q)|^q\,|S(R)|^{q+1} \quad \leq \quad \mathrm{Cov}\left(|S(Q)|^q, (1+|S(R)|)\right) + \mathbb{E}\,|S(Q)|^q\,\mathbb{E}(1+|S(R)|)$$

In the $\lambda-$dependence and the $\omega-$dependence setting, we have respectively :

$$\begin{aligned} \left|\mathrm{Cov}\left(|S(Q)|^q, (1+|S(R)|)\right)\right| &\leq (3q+1)(N\vee 1)^q\,|U|^{q+1}\,\lambda_X(r) \\ \left|\mathrm{Cov}\left(|S(Q)|^q, (1+|S(R)|)\right)\right| &\leq (3q+5)(N\vee 1)^{q+1}\,|U|^{q+2}\,\omega_X(r) \end{aligned}$$

Moreover $\mathbb{E}\,|S(Q)|^q\,\mathbb{E}(1+|S(R)|) \leq C_q^q 2(1\vee C_2)\,|U|^{(q+1)/2}$.
From $\lambda \geq d(q+1)/2$ and $\omega \geq \frac{d}{2}(q+3)/2$ we derive respectively :

$$\begin{aligned} \mathbb{E}\,|S(Q)|^q\,|S(R)| &\leq \left((2q+3)K\xi^{-\lambda}(N\vee 1)^q + 2(1\vee C_2)C_q^q\right)|U|^{(q+1)/2} \\ &\leq \left((2q+5)K\xi^{-\omega}(N\vee 1)^{q+1} + 2(1\vee C_2)C_q^q\right)|U|^{(q+1)/2} \end{aligned}$$

Hence X satisfies condition $\mathbf{C}(q+1)$, if respectively

$$C_{q+1} \geq L^{\frac{1}{q+1}}/(1-\tau(q+1)^{\frac{1}{q+1}} - \sqrt{\xi})$$

for $\xi \in (0, 1/2]$ such that $1 - \tau^{1/(q+1)} - \sqrt{\xi} > 0$ and

$$L = 2(2^{q+1} - 3)\left[(2q+3)K\xi^{-\lambda}(N \vee 1)^q + 2(1 \vee C_2)C_q^q\right]$$

and if

$$C_{q+1} \geqslant L^{\frac{1}{q+1}}/(1 - \tau(q+1)^{\frac{1}{q+1}} - \sqrt{\xi})$$

for $\xi \in (0, 1/2]$ such that $1 - \tau^{\frac{1}{q+1}} - \sqrt{\xi} > 0$ and

$$L = 2(2^{q+1} - 3)\left[(2q+5)K\xi^{-\omega}(N \vee 1)^{q+1} + 2(1 \vee C_2)C_q^q\right]. \square$$

3.4.2 A central limit theorem for weakly dependent random fields

We now prove a central limit theorem for $\eta-$weakly dependent random fields. The proof is based on the Stein method used in Bolthausen [5] to derive the CLT for the partial sums of a stationary strong mixing random field. This proof is adapted from Guyon [25] in order to relax stationarity condition with the moment assumption $\sup_{t \in \mathbb{Z}^d} \|X_t\|_{2+\delta} < \infty$ for $\delta > 0$. We will follow this proof to show that the same holds under λ or ω weakly dependence. The Lemma 3.1 will be extensively used. Let X be a real centered random field and $(D_n)_n$ a sequence of finite subsets of $\mathbb{Z}^d$, $S_n = \sum_{j \in D_n} X_j$, and $\sigma_n^2 = \mathrm{Var}\,(S_n)$. The following assumptions will be used :

($\mathbf{A_1}$) There exists $m > 2$ such that $\sup_{j \in \mathbb{Z}^d} \mathbb{E}\,|X_j|^m < \infty$.

($\mathbf{A_2}$) The random field X satisfies $\lambda_X(r) = \mathcal{O}\left(r^{-\lambda}\right)$, $\lambda > 2d \vee d(m-1)/(m-2)$.

($\mathbf{A_2'}$) The random field X satisfies $\omega_X(r) = \mathcal{O}\left(r^{-\omega}\right)$, $\omega > 3d \vee dm/(m-2)$.

($\mathbf{A_3}$) $\liminf_{n \to \infty} \sigma_n^2/|D_n| > 0$.

Theorem 3.2 *If* ($\mathbf{A_1}$), ($\mathbf{A_2}$) *or* ($\mathbf{A_2'}$) *and* ($\mathbf{A_3}$) *hold, then* $\sigma_n^{-1} S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Proof. We follow here [25] (pages 111-115), quoting that assumption ($\mathbf{A_3}$) implies

$$\sigma_n^{-2} \le C\,|D_n|^{-1} \tag{3.12}$$

for a constant $C > 0$ and for large enough n. In the proof K will denote a generic strictly positive constant which only depends on the random field X.

1) First we show that this is enough to prove the result for bounded random variables. We use a truncation technique. For $T > 0$ let $X_{j,T} = (-T) \vee X_j \wedge T - \mathbb{E}(-T) \vee X_j \wedge T$ if $j \in \mathbb{Z}^d$, $S_{n,T} = \sum_{j \in D_n} X_{j,T}$, and $\sigma_{n,T}^2 = \mathrm{Var}\, S_{n,T}$. We have $\mathbb{E}\left|\sigma^{-1} S_n - \sigma_{n,T}^{-1} S_{n,T}\right|^2 \le 2\sigma_n^{-2}\left(\mathbb{E}\,|S_n - S_{n,T}|^2 + \left|\sigma_n^2 - \sigma_{n,T}^2\right|\right)$. We first bound

$$\begin{aligned}
\sigma_n^{-2}\left|\sigma_n^2 - \sigma_{n,T}^2\right| &\le \sigma_n^{-2} \sum_{i,j \in D_n} \left(|\mathrm{Cov}(X_i - X_{i,T}, X_j)| + |\mathrm{Cov}(X_{i,T}, X_j - X_{j,T})|\right) \\
&\le C \sup_{j \in \mathbb{Z}^d} \sum_{i \in \mathbb{Z}^d} \left(|\mathrm{Cov}(X_i - X_{i,T}, X_j)| + |\mathrm{Cov}(X_{i,T}, X_j - X_{j,T})|\right)
\end{aligned}$$

Note that if in the random field X we replace some variables X_j by $X_{j,T}$ or $X_j - X_{j,T}$, the new random random field denoted Y is allways $\lambda-$weakly dependent or $\omega-$weakly dependent and $\lambda_Y(r) \leqslant \lambda_X(r)$ (resp. $\omega_Y(r) \leqslant \omega_X(r)$). Moreover for a given $j \in \mathbb{Z}^d$

$$\|X_{j,T}\|_m \vee \|X_j - X_{j,T}\|_m \leq 2\,\|X_j\|_m$$

This allows to use the covariance inequality in Lemma 3.1. If $(\mathbf{A_2})$ holds then for a given $i \in \mathbb{Z}^d$ we have :

$$|\mathrm{Cov}(X_i - X_{i,T}, X_j)| + |\mathrm{Cov}(X_{i,T}, X_j - X_{j,T})| \leq K\lambda_X(|i-j|)^{\frac{m-2}{m-1}} \tag{3.13}$$

The same holds if we suppose $(\mathbf{A_2'})$ by replacing the last bound by $K\omega_X(|i-j|)^{\frac{m-2}{m}}$.
Moreover if $1/m + 1/a = 1$ we use Hölder inequality :

$$|\mathrm{Cov}(X_i - X_{i,T}, X_j)| \leq M^{1/m}\left\|X_i \mathbb{1}_{|X_i|>T}\right\|_a \leq MT^{2-m}$$

We bound $|\mathrm{Cov}(X_{i,T}, X_j - X_{j,T})|$ in the same way. We obtain

$$|\mathrm{Cov}(X_i - X_{i,T}, X_j)| + |\mathrm{Cov}(X_{i,T}, X_j - X_{j,T})| \leq KT^{2-m} \tag{3.14}$$

Using (3.13) and (3.14), in the λ-dependent case, we obtain for a given $k \in \mathbb{N}$

$$\sigma_n^{-2}|\sigma_n^2 - \sigma_{n,T}^2| \leq K\Big((2k+1)^d T^{2-m} + \sum_{r>k} r^{d-1}\lambda_X(r)^{\frac{m-2}{m-1}}\Big).$$

The choice of $k = T^\alpha$ with $\alpha \in (0, \frac{m-2}{d})$ gives

$$\lim_{T\to\infty}\limsup_{n\to\infty}\sigma_n^{-2}\left|\sigma_n^2 - \sigma_{n,T}^2\right| = 0 \tag{3.15}$$

Under ω-dependence the same conclusion holds. For $\sigma_n^{-2}\mathbb{E}\left|S_n - S_{n,T}\right|^2$ we derive the same result and finally $\lim_{T\to\infty}\limsup_{n\to\infty}\mathbb{E}\left|\sigma_n^{-1}S_n - \sigma_{n,T}^{-1}S_{n,T}\right|^2 = 0$. With (3.15) and $(\mathbf{A_3})$, we remark that for a given sufficiently large $T > 0$, $\liminf_{n\to\infty}\sigma_{n,T}/|D_n| > 0$. Thus this will be enough to consider the case of bounded variables.

2) We now proceed for bounded random variables. Let $T = \sup_{t\in\mathbb{Z}^d}\|X_t\|_\infty$. Let (m_n) a sequence of positive real numbers such that $\lim_{n\to\infty} m_n = \infty$, $|D_n|^{-1/2}\, m_n^d \to_{n\to\infty} 0$ and

- $|D_n|\,\lambda_X(m_n) \to_{n\to\infty} 0$ if $(\mathbf{A_2})$ holds (this is possible if we set $m_n = |D_n|^\alpha$ with $\alpha \in (\frac{1}{\lambda}, \frac{1}{2d})$ since $\lambda > 2d$).
- $|D_n|\,\omega_X(m_n) \to_{n\to\infty} 0$ if $(\mathbf{A_2'})$ holds (this is possible if we set $m_n = |D_n|^\alpha$ with $\alpha \in (\frac{3}{2\omega}, \frac{1}{2d})$ since $\omega > 3d$).

For $j \in \mathbb{Z}^d$, define $S_{j,n} = \sum_{i\in D_n, d(i,j)\leq m_n} X_i$, $a_n = \sum_{j\in D_n}\mathbb{E}X_j S_{j,n}$, $\overline{S}_n = a_n^{-1/2}S_n$, and $\overline{S}_{j,n} = a_n^{-1/2}S_{j,n}$.
If $(\mathbf{A_2})$ holds we have

$$\sigma_n^{-2}\left|\sigma_n^2 - a_n\right| \leq \sigma^{-2}\sum_{j\in D_n}\sum_{d(i,j)>m_n}|\mathrm{Cov}(X_i, X_j)| \leq 2TC\sum_{k>m_n} k^{d-1}\lambda_X(k)$$

Hence

$$\lim_{n\to\infty} \frac{a_n}{\sigma_n^2} = 1 \tag{3.16}$$

We obtain the same result if $(\mathbf{A_2'})$ holds. Then $\lim_{n\to\infty} \mathbb{E}\left|\sigma_n^{-1} S_n - \overline{S}_n\right|^2 = 0$ and it is enough to show that $\overline{S}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. For this we use the following lemma.

Lemma 3.4 (Stein, [21]) *Let $(\nu_n)_{n\geqslant 1}$ be probability measures over $\mathbb{R}$ with (i) $\sup_n \int x^2 \nu_n(dx) < \infty$, and (ii) $\lim_n \int (i\lambda - x) e^{i\lambda x} \nu_n(dx) = 0$ for all $\lambda \in \mathbb{R}$, then $\nu_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.*

Here (i) is obvious. First observe that from $(\mathbf{A_3})$ and (3.16), there is a constant $C' > 0$ such that $a_n^{-1} \leq C' |D_n|^{-1}$, for n large enough. To prove (ii) use the decomposition $(i\lambda - \overline{S}_n) e^{i\lambda \overline{S}_n} = A_1 - A_2 - A_3$, with $A_1 = i\lambda e^{i\lambda \overline{S}_n}\left(1 - a_n^{-1} \sum_{j\in D_n} X_j S_{j,n}\right)$, $A_2 = a_n^{-1/2} e^{i\lambda \overline{S}_n} \sum_{j\in D_n} X_j \left(1 - i\lambda \overline{S}_{j,n} - e^{-i\lambda \overline{S}_{j,n}}\right)$ and $A_3 = a_n^{-1/2} \sum_{j\in D_n} X_j e^{i\lambda(\overline{S}_n - \overline{S}_{j,n})}$.

Term A_1. If $(\mathbf{A_2})$ holds then

$$\begin{aligned}
\mathbb{E}|A_1|^2 &= \lambda^2 a_n^{-2} \mathrm{Var}\Big(\sum_{j\in D_n} X_j S_{j,n}\Big) \\
&= \lambda^2 a_n^{-2} \sum_{j,j'\in D_n; d(i,j), d(i',j')\leq m_n} \mathrm{Cov}(X_i X_j, X_{i'} X_{j'})
\end{aligned}$$

If $d(j,j') = k \geq 3m_n$, we have $\left|\mathrm{Cov}\left(X_i X_j, X_{i'} X_{j'}\right)\right| \leq K \lambda_X(k - 2m_n)$.
If $d(j,j') \wedge d(j,i) \wedge d(j,i') = k < 3m_n$, we obtain

$$\begin{aligned}
\left|\mathrm{Cov}\left(X_i X_j, X_{i'} X_{j'}\right)\right| &\leq \left|\mathrm{Cov}\left(X_j, X_i X_{i'} X_{j'}\right)\right| \\
&\quad + \left|\mathrm{Cov}\left(X_i, X_j\right)\right| \left|\mathrm{Cov}\left(X_{i'}, X_{j'}\right)\right| \leq K \lambda_X(k)
\end{aligned}$$

thus,

$$\begin{aligned}
\mathbb{E}|A_1|^2 &\leq K a_n^{-2} |D_n| m_n^{2d} \Big(\sum_{k=0}^{3m_n} \lambda_X(k) k^{d-1} + \sum_{k=3m_n}^{\infty} k^{d-1} \lambda_X(k - 2m_n)\Big) \\
&\leq K |D_n|^{-1} m_n^{2d} \to_{n\to\infty} 0
\end{aligned}$$

The same holds if we replace assumption $(\mathbf{A_2})$ by $(\mathbf{A_2'})$.

Term A_2.

$$\begin{aligned}
\mathbb{E}|A_2| &\leq K a_n^{-1/2} \sum_{j\in D_n} \mathbb{E}|X_j| \overline{S}_{j,n}^2 \\
&\leq K a_n^{-3/2} \sum_{j\in D_n} \sum_{d(i,j), d(i',j)\leq m_n} |\mathbb{E} X_{i'} X_i| \\
&\leq K a_n^{-3/2} |D_n| m_n^d \\
&\leqslant K |D_n|^{-1/2} m_n^d \to_{n\to\infty} 0.
\end{aligned}$$

Term A_3. We use weak dependence. We obtain respectively

$$\begin{aligned}
|\mathbb{E}A_3| &\le a_n^{-1/2} \sum_{j\in D_n} \left|\mathbb{E}X_j e^{i\lambda(\overline{S}_n - \overline{S}_{j,n})}\right| \\
&\le a_n^{-1/2} |D_n| \left(1 + |\lambda| (T+1) a_n^{-1/2} |D_n|\right) \lambda_X(m_n) \\
&\le K |D_n| \lambda_X(m_n) \to_{n\to\infty} 0. \\
|\mathbb{E}A_3| &\le a_n^{-1/2} \sum_{j\in D_n} \left|\mathbb{E}X_j e^{i\lambda(\overline{S}_n - \overline{S}_{j,n})}\right| \\
&\le a_n^{-1/2} |D_n| \left(1 + |\lambda| (T+1) a_n^{-1/2} |D_n| + T(1+|D_n|)\right) \omega_X(m_n) \\
&\le K |D_n|^{3/2} \omega_X(m_n) \to_{n\to\infty} 0.
\end{aligned}$$

Then the result follows from Lemma 3.4. □

The following inequality precises the one proved on page 4 of [9].

Lemma 3.5 $(x+y)^{q+\delta} \le x^{q+\delta} + y^{q+\delta} + \left(2^{q+1} - 3\right)\left(x^q y^\delta + y^q x^\delta\right)$ *for* $x, y \ge 0$, $q \in \mathbb{N}$, $q \ge 2$ *and* $0 < \delta \le 1$.

Proof. We write $(x+y)^{q+\delta} = \left(x^q + y^q + \sum_{j=1}^{q-1} \begin{pmatrix} q \\ j \end{pmatrix} x^j y^{q-j}\right)(x+y)^\delta$.

We first note that $(x^q + y^q)(x+y)^\delta \le x^{q+\delta} + y^{q+\delta} + x^\delta y^q + x^q y^\delta$.

For $x \le y$ et $1 \le j \le q-1$, then $x^j y^{q-j}(x+y)^\delta \le 2^\delta x^{j-\delta} x^\delta y^{q-j+\delta} \le 2^\delta x^\delta y^q$.

If $y \le x$, then analogously $x^j y^{q-j}(x+y)^\delta \le 2^\delta y^\delta x^q$. Finally :

$$\begin{aligned}
(x+y)^{q+\delta} &\le x^{q+\delta} + y^{q+\delta} + \left(1 + 2^\delta \textstyle\sum_{j=1}^{q-1} \begin{pmatrix} q \\ j \end{pmatrix}\right)(x^q y^\delta + x^\delta y^q) \\
&\le x^{q+\delta} + y^{q+\delta} + \left(1 + 2^{\delta+1}(2^{q-1} - 1)\right)(x^q y^\delta + x^\delta y^q).
\end{aligned}$$ □

3.4.3 Donsker invariance principle for random fields

We now apply the results of the last subsections to derive a weak invariance principle for λ or ω dependent random fields. We will say that B is a block of $[0,1]^d$ if $B = (s_1, t_1] \times \cdots \times (s_d, t_d]$ with $0 \le s_i < t_i \le 1$ for each $i \in \{1, \ldots, d\}$. If X is a random field, we denote $S_n(B) = \sum_{j\in nB\cap\mathbb{Z}^d} X_j$ and $S_n(t) = S_n((0,t])$ with the notation $S_n(t) = 0$ if $\wedge_{1\leqslant i\leqslant d}\, t_i = 0$.

Under ω dependence, the following assumption will be useful :

$(\mathbf{A_2''})$ X is a $\omega-$weakly dependent random field, $\omega_X(r) = \mathcal{O}\left(r^{-\omega}\right)$, $\omega > 3d \vee d\frac{4m-5}{m-2}$.

We now give the following result :

Theorem 3.3 *Let X be a centered and weakly stationary random field. Assume that, either* $(\mathbf{A_1})$, $(\mathbf{A_2})$ *or* $(\mathbf{A_2''})$ *hold then, for a W a Brownian sheet :*

$$n^{-d/2} S_n(t) \overset{\mathcal{D}([0,1]^d)}{\to} \sigma W(t), \qquad \sigma^2 = \sum_{j\in\mathbb{Z}^d} \mathbb{E}X_0 X_j \geqslant .0$$

Proof. We need to prove both the tightness and the convergence of the finite dimensional distributions of the sequence of processes $(S_n)_n$.
We first prove the convergence of the finite dimensional distributions. Note that it is enough to prove that for all disjoint blocks $B_1, \ldots, B_k$ of $[0,1]^d$:

$$(S_n(B_1), \ldots, S_n(B_k)) \xrightarrow{\mathcal{D}} (W(B_1), \ldots, W(B_k))$$

For simplicity, we prove the result for $k = 2$. Let B, C two disjoint blocks and $\mu, \nu \in \mathbb{R}$. We prove that

$$n^{-d/2}\left(\mu S_n(B) + \nu S_n(C)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(\mu^2\lambda(B) + \nu^2\lambda(C))\right) \tag{3.17}$$

Let $S_n = \mu S_n(B) + \nu S_n(C)$. We can write $S_n = \sum_{j\in\{0,\ldots,n\}^d} Y_{n,j}$ with $Y_{n,j} = \mu X_j$ if $j \in nB$, $Y_{n,j} = \nu X_j$ if $j \in nC$ and $Y_{n,j} = 0$ elsewhere. Let $\sigma_n^2 = \mathrm{Var}\,(S_n)$. We have

$$\sigma_n^2 = \mathrm{Var}\, S_n(B) + \mathrm{Var}\, S_n(C) + 2\mathrm{Cov}\,(S_n(B), S_n(C))$$

– We prove that $n^{-d}\mathrm{Cov}\,(S_n(B), S_n(C)) \to_{n\to\infty} 0$. Indeed as $B \cap C = \emptyset$ we obtain for $0 < \beta < 1$:

$$\begin{aligned} n^{-d}\left|\mathrm{Cov}\,(S_n(B), S_n(C))\right| &\leq n^{-d} \sum_{i\in nB, j\in nC; d(i,nC)\leq n^\beta} |\mathbb{E}X_iX_j| \\ &+ n^{-d} \sum_{i\in nB, j\in nC; d(i,nC)> n^\beta} |\mathbb{E}X_iX_j| \\ &\leq n^{\beta-1}\sum_{j\in\mathbb{Z}^d} |\mathbb{E}X_0X_j| + \sum_{|j|>n^\beta} |\mathbb{E}X_0X_j| \end{aligned}$$

The result follows from Lemma 3.1 (note that in the $\omega-$dependence setting, $\omega > d\frac{4m-5}{m-2} \geq d\frac{m}{m-2}$ and one can apply Lemma (3.1)).

– Recall the notation $S(V) = \sum_{j\in V} X_j$. As $\lim_{|V|\to\infty} \frac{\mathrm{Var}\, S(V)}{|V|} = \sigma^2$:

$$\lim_{n\to\infty} n^{-d}\left(\mathrm{Var}\, S_n(B) + \mathrm{Var}\, S_n(C)\right) = \sigma^2\left(\lambda(B) + \lambda(C)\right)$$

The two previous points imply

$$\lim_{n\to\infty} \frac{\sigma_n^2}{n^d} = \sigma^2\left(\lambda(B) + \lambda(C)\right) \neq 0 \tag{3.18}$$

Now as the triangular array $\{Y_{n,j}/n \in \mathbb{N}, j \in \{0,\ldots,n\}^d\}$ satisfies the same moment condition and the same dependence condition as X, we can prove in the same lines as in the proof of Theorem 3.2 that : $\lim_{n\to\infty} \sigma_n^{-1} S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. With (3.18) we derive (3.17).
We now prove the tightness of the sequence of process S_n. Using $(\mathbf{A_2})$ or $(\mathbf{A_2''})$ and Lemma 3.2 with $q = 2$, there exists $\delta > 0$ and $C > 0$ such that for a block B of $[0,1]^d$,

$$\|S_n(B)\|_{2+\delta} \leq C\left|B \cap \mathbb{Z}^d\right|$$

If the corner points of B are among $\{\frac{i}{n}/i = 0,\ldots,n\}$, we have $\|S_n(B)\|_{2+\delta} \leq C\lambda(B)$ and the condition $(1+\delta/2, 1+\delta/2)$ of Bickel and Wichura [4] holds. Then from Theorem 3 in [4] and the remark below it, we derive the tightness of the sequence $(S_n)_n$. □

Bibliographie

[1] Andrews, D. Non strong mixing autoregressive processes, Journal of Applied Probability 21, 930-934, (1984)

[2] Bardet, J. M., Doukhan, P., León, J. R. Uniform limit theorems for the integrated periodogram of weakly dependent time series and their applications to Whittle's estimate. To appear in J.T.S.A. `hal.archives-ouvertes.fr/hal-00126489/en/`, (2007)

[3] Bardet, J. M., Doukhan, P., León, J. R. A uniform central limit theorem for the periodogram and its applications to Whittle parametric estimation for weakly dependent time series. ESAIM P & S 12, 154Ű172, (2008)

[4] Bickel, P. J., Wichura, M. J. Convergence criteria for multiparameter stochastic processes and some applications. (English) Ann. Math. Stat. 42, 1656-1670, (1971)

[5] Bolthausen, E. On the C.L.T for stationary mixing random fields, Ann. Probab. 10-4, 1047-1050, (1982)

[6] Bradley, R. Introduction to Strong Mixing Conditions Volumes 1,2 and 3. Kendricks Press, (2007)

[7] Bulinski, A. V., Shashkin, A. P. Rates in the central limit theorem for weakly dependent random variables, J. Math. Sci., 122, 3343-3358, (2004)

[8] Bulinski, A. V., Shashkin, A.P. Strong Invariance Principle for Dependent Multi-indexed Random Variables, Doklady Mathematics 72-1, 503-506, (2005)

[9] Bulinski, A. V., Shashkin, A. Strong invariance principle for dependent random fields, IMS Lecture Notes–Monograph Series 48, 128-143, Providence, (2006)

[10] Dedecker, J., Doukhan, P. A new covariance inequality and applications, Stochastic Processes and their Applications 106, 63-80, (2003)

[11] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., Prieur, C. Weak dependence, examples and applications, Lecture Notes in Statistics 190, (2007)

[12] Dedecker, J., Prieur, C. New dependence coefficients. Examples and applications to statistics. Probability Theory and Related Fields 132, 203-236, (2005)

[13] Doukhan, P. Mixing : Properties and Examples, LNS 85, Springer Verlag, New-York, (1994)

[14] Doukhan, P. Models inequalities and limit theorems for stationary sequences, in Theory and applications of long range dependence (Doukhan *et alii* ed.) Birkhäuser, 43-101, (2003)

[15] Doukhan, P., Lang, G. (2002) Rates in the empirical central limit theorem for stationary weakly dependent random fields, Statistical Inference Stochastic Processes 5, 199-228, (2002)

[16] Doukhan, P., Louhichi, S. A new weak dependence condition and applications to moment inequalities, Stochastic Processes and their Applications, Vol. 84, 313-342, (1999)

[17] Doukhan, P., Madre, H., Rosenbaum, M. Weak dependence for infinite ARCH-type bilinear models, Statistics 41, 31-45, (2007)

[18] Doukhan, P., Teyssière, G., Winant, P. A LARCH(∞) vector valued process, in Lecture Note in Statistics 187, Dependence in Probability and statistics (Bertail, P., Doukhan, P. and Soulier, P. editors), 245-258, (2006)

[19] Doukhan, P., Truquet, L. Weakly dependent random fields with infinite memory. Alea 2, 111-132, (2007).

[20] Doukhan, P., Wintenberger, O. A central limit theorem under non causal weak dependence and sharp moment assumptions, Prob. Math. Stat. 27, 45-73, (2007)

[21] Doukhan, P., Wintenberger, O. Weakly dependent chains with infinite memory, to appear in Stoch. Proc. And Applic., (2008)

[22] Engle, R. F. Autoregressive conditional heteroskedasticity with estimate of the variance in the U.K. inflation. Econometrica 15, 286-301, (1982)

[23] Giraitis, L., Kokoszka, P., Leipus, R. Stationary $ARCH-$models : dependence structure and central limit theorem. Econometric Theory, 16, 3-22, (2000)

[24] Giraitis, L., Surgailis, D. ARCH-type bilinear models with double long memory. Stoch. Proc. and their Applic. 100, 275-300, (2002)

[25] Guyon, X. Random Fields on a Network, Springer-Verlag, New-York, (1991)

[26] Petrov, V. Limit theorems of probability theory. Clarendon Press, Oxford, (1995)

[27] Rio, E. Théorie asymptotique pour des processus aléatoires faiblement dépendants. SMAI, Mathématiques et Applications 31, Springer, (2000)

[28] Stein, C. A bound for the error in normal approximation of a sum of dependent random variables. Proc. Berkeley Sympos., M.S.P.2, 583-603, (1972)

Chapitre 4

Strong invariance principle for a new class of weakly dependent random fields

Abstract

We prove a strong invariance principle for some (non-causal) spatial Bernoulli shifts with independent and identically distributed (i.i.d.) inputs. Several examples of such random fields are given in Doukhan and Truquet [13], where the authors use a fixed point theorem to solve autoregressive equations of the form

$$X_t = F\left((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_t\right). \tag{4.1}$$

where ξ is an i.i.d. random field. Classical mixing conditions are hard or impossible to obtain for such models but weak dependence conditions of Doukhan and Louhichi (1999) are satisfied. The main tools to derive the strong approximation are the Berkes and Morrow (1981) blocking technique, moment inequalities for weakly dependent random fields and an approximation by $m-$dependent random fields. Moreover, only a Riemanian rate for the covariances is required to establish a strong invariance principle which notably extends previous results even for causal fields.

4.1 Introduction

Bernoulli shifts often refer to time series such as ARMA models or ARCH processes. Their extension to the multiparameter case is well known only for the ARMA random fields which have been extensively studied (see [16]). In a recent paper of Doukhan and Truquet [13], a new class of random fields has been introduced, extending linear random fields. Their construction is based on the resolution of autoregressive equations of the form

$$X_t = F\left((X_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}};\xi_t\right). \tag{4.2}$$

Under contraction assumptions on the function F, it is proved that there exists a unique solution written as a spatial Bernoulli shift of the form :

$$X_t = H\left((\xi_{t-j})_{j\in\mathbb{Z}^d}\right), \quad t \in \mathbb{Z}^d, \tag{4.3}$$

where H is a measurable function. The random field ξ can be supposed only stationary. The construction of such spatial processes is done through moments conditions, using a fixed point theorem. An approximate simulation of the solution of equation (4.2) can be done using a recursive procedure (see section 2.4 in [13]).

Classical cross moments and others statistics are difficult to obtain for a model defined by a general equation of the form (4.2) except in the case of linear fields or with the use of a convenient notion of causality, which also relaxes the moment conditions on the inputs. In this last case, one obtain a generalization of the classical autoregressive time series such as the classical ARCH models, and a non linear extension of the quarter plane AR (see [17]).

For $X = (X_t,\, t \in \mathbb{Z}^d)$ a spatial Bernoulli shift satisfying (4.3), denote the partial sum

$$S_N = \sum_{0<j\leq N} X_j$$

where for $i, j \in \mathbb{Z}^d$, we use the notations $i \leq j$ if $i_s \leq j_s$ for all $s = 1, \ldots, d$ and $i < j$ if $i_s < j_s$ for all $s = 1, \ldots, d$, $N = (N_1, \ldots, N_d) \in G_\tau$ and for $\tau \in (0, 1)$,

$$G_\tau = \bigcap_{s=1}^{d} \Big\{ j = (j_1, \ldots, j_d) \in \mathbb{N}^d,\, j_s \geq \big(\prod_{t\neq s} j_t\big)^\tau \Big\}.$$

The goal of this paper is to establish a strong approximation for S_N in the particular case of spatial Bernoulli shifts with i.i.d. inputs, that is an approximation of the form,

$$S_N - W_N = O\left([N]^{1/2-\varepsilon}\right) \quad a.s. \tag{4.4}$$

when $N \in G_\tau \to \infty$ (that means $N_i \to \infty$ for all $1 \leq i \leq d$) where

- $[N] = \prod_{s=1}^{d} N_s$;
- W is a multi-parameter Wiener process ;
- $\varepsilon > 0$ is not depending on N.

The first result of strong approximation for random fields has been given in Berkes and Morrow (1981) [4] in the case of strong mixing where the authors have used the reconstruction method of Berkes and Philipp (1979) [5] to construct an approximation in the multivariate case. More recently, Balan (2005) [3] have proved a strong invariance principle for associated random fields using the quantile transform of Csörgo and Révész [8]. A generalization to a more general weak dependence condition is given in Bulinski and Shashkin (2005) [6], with the same method.

Relating to the limit theory of random fields defined by (4.3), classical mixing conditions for models defined by equation (4.2) are very difficult to be obtained in practice. Moreover there are some simple examples of causal sequences which are not strong mixing : the AR model proposed by Andrews [1] or a ARCH model provided in [12]. On the other hand, the weak dependence introduced by Doukhan and Louhichi [11] is an appropriate property to study the limit theory of non causal Bernoulli shifts, with independent inputs [10] as well as for dependent ones [14]. General limit theorems such as moments inequalities and weak invariance principle have also been proved in [12] for more general models than (4.3).
Nevertheless, in the time series case, powerful methods have been recently used for the strong approximation of partial sums of Bernoulli shifts with i.i.d inputs. Wu (2007) [20] have used a martingale approximation for general causal Bernoulli shifts. Approximation by $m-$dependent sequences is used by Aue and al. (2006) [2] for augmented GARCH sequences. Liu and Lin (2008) [18] have taken advantage of an approximation by $m-$dependent sequences in the general case, the result of Wu (2007) is improved and some optimal rates have been obtained.
In this paper, we will use an approximation by $m-$dependent random fields to construct the Brownian sheet. In the time series case, moment inequalities for Bernoulli shifts used in [20] or [18] are derived from martingale difference decompositions and the Burkholder inequality. Since we have not found a similar decomposition in the case of random fields, we will use a general moment inequality given in [12] for weakly dependent random fields which is in fact an adaptation from that of Bulinski and Shashkin [6].
In contrast with time series, there is a technical difficulty with the random fields because as it has been pointed in [4] the strong approximation cannot hold near the coordinate axes due to the irregular behaviour of the variance in this case. This leads to the introduction of an "equilibrate" subset (such as G_τ) for the convergence of the partial sums.
The paper is organized as follows. Some bounds for the covariances of spatial Bernoulli shifts that will be useful for the approximation of the random field by $m-$dependent fields are provided in next Section 4.2. Then, we recall the notion of weak dependence of Doukhan and Louhichi for random fields and establish some required moment inequalities in Section 4.3. Finally, Section 4.4 is devoted to the strong invariance principle and examples. The proof of this result is postponed to the last section of the paper, Section 4.5.

4.2 Covariances estimates for Bernoulli shifts

Let K be a positive integer. We denote by $\|\cdot\|$ the Euclidian norm on $\mathbb{R}^K$ and the same notation will be used for the associated norm on $\mathcal{M}_{K,K}$ the space of $K \times K$ real square matrices. Moreover, $\|\cdot\|_\infty$ will denote the suppremum norm on $\mathbb{Z}^d$.

If X, Y are two random vectors with values in $\mathbb{R}^K$, we denote by $\Gamma_{X,Y}$ the covariance operator with values in $\mathcal{M}_{K,K}$:

$$\Gamma(X,Y) = \mathbb{E}\left((Y-\mathbb{E}Y)(X-\mathbb{E}X)'\right).$$

Now for $h \geq 1$, we denote

$$\|X\|_h = \left[\mathbb{E}\,\|X\|^h\right]^{1/h}.$$

Note that we have the following Cauchy-Shwarz inequality :

$$\|\Gamma(X,Y)\| \leq \|X-\mathbb{E}X\|_2\,\|Y-\mathbb{E}Y\|_2\,. \tag{4.5}$$

Let X be a centered random field with values in $\mathbb{R}^K$. We suppose that X is a Bernoulli shift i.e

$$X_t = H\left((\xi_{t-j})_{j\in\mathbb{Z}^d\setminus\{0\}}\right) \tag{4.6}$$

where ξ is an i.i.d random field with value in a measurable space E' and $H : E'^{\mathbb{Z}^d} \to \mathbb{R}^K$ is a measurable function. In this section, we suppose that $\mathbb{E}\,\|X_0\|^2 < \infty$. For $t \in \mathbb{Z}^d$ and $l \in \mathbb{N}^*$, we denote the $\sigma-$algebra

$$\mathcal{F}_{t,l} = \sigma\left(\xi_{t-j}/\,\|j\|_\infty < l\right).$$

We also define the following $l-$dependent random field $X_{t,l} = \mathbb{E}_{\mathcal{F}_{t,l}}X_t$ for each $t\in\mathbb{Z}^d$. If l is an odd integer, we set $X_{t,l/2} = X_{t,(l+1)/2}$. In the case $l=0$, we set $X_{t,l} = X_t$, $\forall t\in\mathbb{Z}^d$.
The following proposition provides inequalities satisfied by the covariances of a Bernoulli shift and will be useful to derive bounds for the approximation of X by the $l-$dependent random field $X_{.,l}$. We also denote by $Y_{t,l} = X_t - X_{t,l}$ and

$$p(l) = \|Y_{0,l}\|_2\,.$$

Remark that by usual properties of the conditional expectation, $(p(l))_l$ is non increasing.

Proposition 4.1 *Let $i,j\in\mathbb{Z}^d$ and $l\in\mathbb{N}^*$. We set $\widetilde{l} = \frac{\|i-j\|_\infty}{2}$.*

1. $\|\Gamma\left(X_i,X_j\right)\| \leq 2\,\|X_0\|_2\,p(\widetilde{l})\,\mathbb{1}_{\widetilde{l}\neq 0} + Var\left(X_0\right)\mathbb{1}_{\widetilde{l}=0}$;
2. $\|\Gamma\left(Y_{i,l},Y_{j,l}\right)\| \leq 2p(l)p(\widetilde{l}\vee l)\,\mathbb{1}_{\widetilde{l}\neq 0} + p(l)^2\,\mathbb{1}_{\widetilde{l}=0}$;
3. $\|\Gamma\left(X_i,X_j\right) - \Gamma\left(X_{i,l},X_{j,l}\right)\| \leq \left(2\,\|X_0\|_2\,p(\widetilde{l}\vee l) + 2p(\widetilde{l})p(l)\right)\mathbb{1}_{\widetilde{l}\neq 0} + \|X_0\|_2\,p(l)\,\mathbb{1}_{\widetilde{l}=0}$.

Proof of Proposition 4.1 We will use the following fact : if $\widetilde{l}\neq 0$, then the two $\sigma-$algebras $\mathcal{F}_{i,\widetilde{l}}$ and $\mathcal{F}_{j,\widetilde{l}}$ are independent. In this case, we have $\Gamma\left(X_{i,\widetilde{l}},X_{j,\widetilde{l}}\right) = 0$.

1. The result is obvious if $\widetilde{l}=0$. If $\widetilde{l}\neq 0$, with the Cauchy-Schwarz inequality
$$\begin{aligned}\left\|\Gamma\left(X_i,X_j\right)\right\| &\le \left\|\Gamma\left(Y_{i,\widetilde{l}},X_j\right)\right\|+\left\|\Gamma\left(X_{i,\widetilde{l}},Y_{j,\widetilde{l}}\right)\right\|\\ &\le p(\widetilde{l})\left\|X_0\right\|_2+p(\widetilde{l})\left\|X_{0,\widetilde{l}}\right\|_2\\ &\le 2\left\|X_0\right\|_2 p(\widetilde{l}).\end{aligned}$$
Note that the last inequality follows from Jensen inequality.
2. We apply the first point to the special Bernoulli shift $(Y_{t,l})_{t\in\mathbb{Z}^d}$. If $\widetilde{l}\neq 0$, we obtain
$$\left\|\Gamma\left(Y_{i,l},Y_{j,l}\right)\right\|\le 2\left\|Y_{0,l}\right\|_2\left\|Y_{0,l}-\mathbb{E}_{\mathcal{F}_{0,\widetilde{l}}}(Y_{0,l})\right\|_2.$$
From properties of the conditionnal expectation $Y_{0,l}-\mathbb{E}_{\mathcal{F}_{0,\widetilde{l}}}(Y_{0,l})=Y_{0,\widetilde{l}\vee l}$ and the result follows. If now $\widetilde{l}=0$, we remark that $\left\|\Gamma\left(Y_{i,l},Y_{i,l}\right)\right\|\le p(l)^2$.
3. if $\widetilde{l}\neq 0$, we have :
$$\left\|\Gamma\left(X_i,X_j\right)-\Gamma\left(X_{i,l},X_{j,l}\right)\right\| \le \left\|\Gamma\left(Y_{i,l},X_j\right)\right\|+\left\|\Gamma\left(X_{i,l},Y_{j,l}\right)\right\|.$$
But
$$\left\|\Gamma\left(Y_{i,l},X_j\right)\right\| \le \left\|\Gamma\left(Y_{i,l}-\mathbb{E}_{\mathcal{F}_{i,\widetilde{l}}}Y_{i,l},X_j\right)\right\|+\left\|\Gamma\left(\mathbb{E}_{\mathcal{F}_{i,\widetilde{l}}}Y_{i,l},Y_{j,\widetilde{l}}\right)\right\|.$$
As $\left\|\mathbb{E}_{\mathcal{F}_{i,\widetilde{l}}}Y_{i,l}\right\|_2\le p(l)$ and $\left\|Y_{i,l}-\mathbb{E}_{\mathcal{F}_{i,\widetilde{l}}}(Y_{i,l})\right\|_2=\left\|Y_{i,\widetilde{l}\vee l}\right\|_2\le p(\widetilde{l}\vee l)$, the Cauchy-Schwarz inequality provides
$$\left\|\Gamma\left(Y_{i,\widetilde{l}},X_j\right)\right\|\le\left\|X_0\right\|_2 p(\widetilde{l}\vee l)+p(\widetilde{l})p(l).$$
Moreover,
$$\begin{aligned}\left\|\Gamma\left(X_{i,l},Y_{j,l}\right)\right\| &\le \left\|X_{i,l}-\mathbb{E}_{\mathcal{F}_{i,\widetilde{l}}}(X_{i,l})\right\|_2\left\|Y_{j,l}\right\|_2+\left\|Y_{j,l}-\mathbb{E}_{\mathcal{F}_{j,\widetilde{l}}}(Y_{j,l})\right\|_2\left\|\mathbb{E}_{\mathcal{F}_{j,\widetilde{l}}}X_{i,l}\right\|_2\\ &\le \left\|X_{i,l}-\mathbb{E}_{\mathcal{F}_{i,\widetilde{l}}}(X_{i,l})\right\|_2 p(l)+\left\|X_0\right\|_2 p(\widetilde{l}\vee l).\end{aligned}$$
As $\left\|X_{i,l}-\mathbb{E}_{\mathcal{F}_{i,\widetilde{l}}}(X_{i,l})\right\|_2^2\le p(\widetilde{l})^2$ by Jensen Inequality, we obtain
$$\left\|\Gamma\left(X_{i,l},Y_{j,l}\right)\right\|\le\left\|X_0\right\|_2 p(\widetilde{l}\vee l)+p(\widetilde{l})p(l)$$
and the result follows.
For the case $\widetilde{l}=0$ we observe that
$$\left\|\Gamma\left(X_i,X_j\right)-\Gamma\left(X_{i,l},X_{j,l}\right)\right\|\le\left\|X_0\right\|_2 p(l)\le\left\|X_0\right\|_2 p(l).\square$$

We deduce the following corollary.

Corollary 4.1 *Suppose that $p(l)=O\left(l^{-\eta}\right)$ with $\eta>d$.*
1) $\sum_{j\in\mathbb{Z}^d}\left\|\Gamma\left(X_0,X_j\right)\right\|<\infty$.
2) $\sum_{j\in\mathbb{Z}^d}\left\|\Gamma\left(Y_{0,l},Y_{j,l}\right)\right\|=O\left(l^{d-2\eta}\right)$.
3) $\sum_{j\in\mathbb{Z}^d}\left\|\Gamma\left(X_0,X_j\right)-\Gamma\left(X_{0,l},X_{j,l}\right)\right\|=O\left(l^{d-\eta}\right)$.

4.3 Moment inequalities for weakly dependent fields

The goal of this section is to establish a general moment inequality for weakly dependent random fields that can be applied to Bernoulli shifts. A general notion of weak dependence has been introduced by Doukhan and Louhichi (1999). We recall here the definition of the $\eta-$weak dependence for random fields. First, for a function $h : \left(\mathbb{R}^K\right)^w \to \mathbb{R}$ with $K, w \in \mathbb{N}^*$, define

$$\operatorname{Lip} h = \sup_{(x_1,\dots,x_w)\neq(y_1,\dots,y_w)} \frac{|h(x_1,\dots,x_w) - h(y_1,\dots,y_w)|}{\|x_1 - y_1\| + \cdots + \|x_w - y_w\|},$$

where a norm $\|\cdot\|$ is given on $\mathbb{R}^K$. Then,

Definition 4.1 *A $E = \mathbb{R}^K-$valued random field $(X_t)_{t\in\mathbb{Z}^d}$ is weakly dependent if there exists a sequence $(\eta_X(r))_{r\in\mathbb{N}}$ with $\lim_{r\to\infty}\eta_X(r) = 0$ such that for all $u \in \mathbb{N}^*$, $v \in \mathbb{N}^*$ and functions $f : \left(\mathbb{R}^K\right)^u \to \mathbb{R}$ and $g : \left(\mathbb{R}^K\right)^v \to \mathbb{R}$ satisfying $\|f\|_\infty \leqslant 1$, $\|g\|_\infty \leqslant 1$, $Lip\, f < \infty$ and $Lip\, g < \infty$,*

$$|Cov\left(f(X_{s_1},\dots,X_{s_u}),\, g\left(X_{t_1},\dots,X_{t_v}\right)\right)| \leqslant \left(uLip\, f + vLip\, g\right)\eta_X(r),$$

where indices $s_1,\dots,s_u$, $t_1,\dots,t_v \in \mathbb{Z}^d$ are such that $\|s_m - t_l\|_\infty \geq r$ for $1 \leq m \leq u$ and $1 \leq l \leq v$.

A special case of $\eta-$weakly dependent random fields are the Bernoulli shifts. This is proved in [10] (see paragraph 2.2.3 of that paper). Moreover, inspection of the proof given in [10] shows that $\eta_X(r) = 2\delta(r/2)$ where

$$\delta(r) = \left\|H(\left(\xi_j\right)_j) - H\left(\left(\xi_j \mathbb{1}_{\|j\|_\infty<r}\right)_j\right)\right\|_1.$$

A straightforward modification of that proof shows that one can take :

$$\eta_X(r) = 2p(r/2). \tag{4.7}$$

In the sequel a set $U \subset \mathbb{Z}^d$ is called a block if it can be written as $\prod_{s=1}^d (a_s, b_s]$ for $a_s, b_s \in \mathbb{Z}$ with $a_s < b_s$ and denote $|U| = \prod_{s=1}^d (b_s - a_s)$. For a random field X, let :

$$S(U) = \sum_{j\in U} X_j \quad \text{and} \quad M(U) = \sup_{W \text{ block } \subset U} \{|S(W)|\}.$$

We will use the following assumption :

(A1) $X = (X_t)_{t\in\mathbb{Z}^d}$ is a spatial Bernoulli shift with values in $\mathbb{R}^K$, $K \in \mathbb{N}^*$, satisfying (4.3) where $(\xi_t)_{t\in\mathbb{Z}^d}$ are i.i.d. random vectors, and there exist $h > 2$ and $\widetilde{C} > 0$ such that $\mathbb{E}\left\|X_0\right\|^h < \infty$ and for all $r \in \mathbb{N}$, $p(r) \leq \widetilde{C}(r+1)^{-\eta}$ where $\eta > 2d\,\frac{h-1}{h-2}$.

The following theorem provides a moment inequality and a maximal inequality for $\eta-$weakly dependent random fields.

Theorem 4.1 *Suppose that the random field X satisfies* **(A1)**. *Then there exists $\delta \in (0,1)$ and a constant $C > 0$ such that for all block $U \subset \mathbb{Z}^d$:*

$$\|S(U)\|_{2+\delta} \leq C\,|U|^{1/2}, \qquad \|M(U)\|_{2+\delta} \leq AC\,|U|^{1/2}, \tag{4.8}$$

where $A = 5^{\frac{d}{2+\delta}}(1 - 2^{\delta/(4+2\delta)})^{-d}$.

Proof of Theorem 7.8 Since the random field X is a square integrable Bernoulli shift, then it is $\eta-$weakly dependent with dependence coefficients η_X satisfying inequality (4.7). By assumption **(A1)**, we have $\eta_X(r) \leq 2^{1+\eta}\widetilde{C}(r+1)^{-\eta}$. One can observe that each coordinate $(X_t(y))_{t\in\mathbb{Z}^2}$ is also a $\eta-$weakly dependent random field with
$\eta_{X(y)}(r) \leq 2^{1+\eta}\widetilde{C}(r+1)^{-\eta}$, $\eta > 2d\frac{h-1}{h-2}$. Then, using the point 1 of Lemma 4.1, it is possible to apply the moment inequality of Lemma 2 in [12] to each coordinate of X. One can deduce that there exists $\delta > 0$ and $C > 0$ such that for all block $U \subset \mathbb{Z}^d$:

$$\|S(U)\|_{2+\delta} \leq C\,|U|^{1/2}.$$

Lemma 2 in [12] shows that δ is chosen such that $2 + \delta < h$ and satisfies :

$$\eta \geq \frac{d}{2}\,\frac{h(4-\delta) + (\delta - 2)(2+\delta)}{h - 2 - \delta}.$$

The first inequality of Theorem 1 follows and the second inequality is a consequence of the first one and the Móricz theorem [19].□

Remark The constant C can be chosen such that $C \geq \sum_{y=1}^{K} C_y$ where constants C_y are given at the end of the proof of lemma 2 in [12] and are such that :

$$C_y = \left(50\,\mathbb{E}\,|X_0(y)|^h + 210\,K\,\tau_2^{-\eta} + 80\,(C_{2,y} \vee 1)C_{2,y}^2\right)^{\frac{1}{2+\delta}}\left(1 - \tau_1^{\frac{1}{2+\delta}} - \tau_2^{1/2}\right)^{-1}, \tag{4.9}$$

where τ_1 and τ_2 are positive constants which do not depend of the random field X and $C_{2,y} = \left(\sum_{j\in\mathbb{Z}^d} |\mathrm{Cov}\,(X_0(y), X_j(y))|\right)^{1/2}$ with $X_j = (X_j(1), \cdots, X_j(K))$ for all $j \in \mathbb{Z}^d$.

In the next proposition we derive a moment inequality with the same δ and C as in Lemma 7.8 for all the random fields of interest $X_{.,n}$, $n \in \mathbb{N}^* \cup \{\infty\}$.

Corollary 4.2 *Under assumptions* **(A1)**, *there exist $\delta \in (0,1)$ and $C > 0$ such that for all $n \in \mathbb{N}^* \cup \{\infty\}$ and all block $U \subset \mathbb{Z}^d$:*

$$\|S_n(U)\|_{2+\delta} \leq C\,|U|^{1/2}, \quad \text{where } S_n(U) = \textstyle\sum_{j\in U} X_{j,n}.$$

Proof of corollary 4.2 Note that for all $n \in \mathbb{N}^* \cup \{\infty\}$, the random field $X_{.,n}$ is $\eta-$weakly dependent since it is $m-$dependent. We want to show that Theorem 7.8 applies to all the random fields $X_{.,n}$, $n \in \mathbb{N}$ with the same numbers δ and C. In particular, we need a bound, independent of the integer n, for the constant C of the previous remark using (4.9). For this, it is enough to prove the following three points :

– $\sup_{n \in \mathbb{N}^* \cup \{\infty\}} \|X_{0,n}\|_h < \infty$;
– $\sup_{n \in \mathbb{N}^* \cup \{\infty\}} \sum_{j \in \mathbb{Z}^d} |\mathrm{Cov}\,(X_{0,n}, X_{j,n})| < \infty$;
– $\sup_{n \in \mathbb{N}^* \cup \{\infty\}} \eta_{X_{.,n}}(r) \leq 2^{1+\eta}\widetilde{C}(1+r)^{-\eta}$.

By Jensen inequality, $\sup_{n \in \mathbb{N}^* \cup \{\infty\}} \|X_{0,n}\|_h \leq \|X_0\|_h$ and this proves the first point.
The second point is a straightforward consequence of Proposition 4.1. Indeed, there exist $C_1 > 0$ and $C_2 > 0$ not depending on n such that

$$\begin{aligned} \|\Gamma\,(X_{0,n}, X_{j,n})\| &\leq \|\Gamma\,(X_{0,n}, X_{j,n}) - \Gamma\,(X_0, X_j)\| + \|\Gamma\,(X_0, X_j)\| \\ &\leq C_1\, p(\|j\|_\infty/2)\mathbb{1}_{j\neq 0} + C_2\,\mathbb{1}_{j=0}. \end{aligned}$$

Hence the result follows.
To show the last point, consider $u \in \mathbb{N}^*$, $v \in \mathbb{N}^*$ and functions $f : \left(\mathbb{R}^K\right)^u \to \mathbb{R}$ and $g : \left(\mathbb{R}^K\right)^v \to \mathbb{R}$ satisfying $\|f\|_\infty \leqslant 1$, $\|g\|_\infty \leqslant 1$, $\mathrm{Lip}\, f < \infty$ and $\mathrm{Lip}\, g < \infty$. Define also,

- $f_n = f\,(X_{s_1,n}, \dots, X_{s_u,n})\,;\qquad g_n = g\,(X_{t_1,n}, \dots, X_{t_v,n})$
- $X_{t,n,r} = \mathbb{E}_{\mathcal{F}_{t,r}} X_{t,n}, \quad \forall t \in \mathbb{Z}^d$
- $f_{n,r} = f\,(X_{s_1,n,r}, \dots, X_{s_u,n,r})\,;\qquad g_{n,r} = g\,(X_{t_1,n,r}, \dots, X_{t_v,n,r})$

We have

$$\begin{aligned} |\mathrm{Cov}\,(f_n, g_n)| &\leq \left|\mathrm{Cov}\left(f_n - f_{n,\frac{r}{2}}, g_n\right)\right| + \left|\mathrm{Cov}\left(f_{n,\frac{r}{2}}, g_n - g_{n,\frac{r}{2}}\right)\right| \\ &\leq 2\,\mathrm{Lip}\, f \sum_{s=1}^{u} \left\|X_{i_s,n,\frac{r}{2}} - X_{i_s,n}\right\|_2 + 2\,\mathrm{Lip}\, g \sum_{s=1}^{v} \left\|X_{j_s,n,\frac{r}{2}} - X_{j_s,n}\right\|_2 \\ &\leq 2\,(u\mathrm{Lip}\, f + v\mathrm{Lip}\, g) \left\|X_{0,n,\frac{r}{2}} - X_{0,n}\right\|_2 \\ &\leq 2\,(u\mathrm{Lip}\, f + v\mathrm{Lip}\, g)\, p(r/2). \end{aligned}$$

Since by Assumption **(A1)**, $2\,p(r/2) \leq 2^{1+\eta}\widetilde{C}(1+r)^{-\eta}$, we have proved the third point and this leads to the result. □

4.4 The strong invariance principle for spatial Bernoulli shifts

Now we give the main result of this paper : a strong approximation of partial sums of spatial Bernoulli shifts satisfying Assumption **(A1)** by a Wiener process. However, another (weak and only technical) assumption has to be added :

(A2) $\Gamma = \sum_{j\in\mathbb{Z}^d} \Gamma(X_0, X_j)$ is a positive definite matrix.

Theorem 4.2 *Under Assumptions* **(A1)** *and* **(A2)** *and for any* $\tau \in (0,1)$*, there exists* $\varepsilon > 0$ *and* X *can be redefined without changing its distribution on a richer probability space together with a* $d-$*parameter Wiener process* $W = \{W_t, t \in [0,\infty)^d\}$ *such that (4.4) holds.*

Examples : General examples are provided by models defined in [13] and satisfying the equation (4.2). Let $\xi = (\xi_t)_{t\in\mathbb{Z}^d}$ be an i.i.d. random field with values in E' (usually $E' = \mathbb{R}^{K'}$ for some $K' \in \mathbb{N}^*$ but E' can also be a denumerable tensor product of such sets). Let $E = \mathbb{R}^K$ with $K \in \mathbb{N}^*$ and $F : \big(E^{(\mathbb{Z}^d\setminus\{0\})} \times E', \mathcal{B}(E^{\mathbb{Z}^d\setminus\{0\}}) \otimes \mathcal{B}(E')\big) \to \big(E, \mathcal{B}(E)\big)$ be a measurable function (here, if V is a vector space and B an arbitrary set then $V^{(B)} \subset V^B$ denotes the set of $v = (v_b)_{b\in B}$ such that there exists a finite subset $B_1 \subset B$ such that $v_b = 0$ for each $b \notin B_1$). Define also both these assumptions :

- **(H1)** There exists $h > 2$ such that $\|F(0;\xi_0)\|_h < \infty$.
- **(H2)** There exists $(a_j)_{j\in\mathbb{Z}^d\setminus\{0\}}$ such that $a_j \geqslant 0$ for all $j \in \mathbb{Z}^d \setminus \{0\}$ and $a = \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j < 1$, satisfying for all $\forall z, z' \in E^{(\mathbb{Z}^d\setminus\{0\})}$,

$$\|F(z;\xi_0) - F(z';\xi_0)\| \ \le \sum_{j\in\mathbb{Z}^d\setminus\{0\}} a_j\|z_j - z'_j\|, \quad a.s. \tag{4.10}$$

Then,

Proposition 4.2 *Assume that* **(H1)** *and* **(H2)** *hold. Then there exists a unique stationary solution* $X = (X_t)_{t\in\mathbb{Z}^d}$ *of equation (4.2) such that* $\mathbb{E}\|X_0\|^h < \infty$ *which can be written* $X_t = H\big((\xi_{t-j})_{j\in\mathbb{Z}^d}\big)$ *for all* $t \in \mathbb{Z}^d$*, where* H *is a measurable function.*

This proposition follows from Theorem 3 in [13]. A similar existence theorem with a convenient notion of causality is also provided in [13], where the assumption **(H2)** is weakened.
Under **(H1)** and **(H2)**, it is possible to check the Assumption **(A1)** in particular cases. Indeed, for $p \in \mathbb{N}^*$ and $t \in \mathbb{Z}^d$, consider the sequence of random fields $\big(X_{p,t}(n)\big)_{n\in\mathbb{N}}$ defined by :

$$X_{p,t}(0) = 0, \quad \text{and} \quad X_{p,t}(n+1) = F\big((X_{p,t-j}(n)\mathbb{1}_{0<\|j\|_\infty\le p})_j;\xi_t\big) \quad \text{for} \quad n \in \mathbb{N}.$$

From Lemma 6 in [13], we have the following bound :

$$\|X_0 - X_{p,0}(n)\|_h \le \|X_0\|_h\Big(a^n + \frac{1}{1-a}\sum_{\|j\|_\infty>p} a_j\Big).$$

Now since the random variable $X_{p,0}(n)$ is measurable with respect to $\mathcal{F}_{0,np+1}$, we have $\mathbb{E}_{\mathcal{F}_{0,l}} X_{p,0}(n) = X_{p,0}(n)$ for $l \geq np+1$. For $p \in \mathbb{N}^*$, we set $n = E(\frac{l-1}{p})$. Then, $np+1 \leq l$ and we obtain

$$p(l) \leq 2\,\|X_0 - X_{p,0}(n)\|_2 \leq 2\,\|X_0\|_h \Big(a^n + \frac{1}{1-a} \sum_{\|j\|_\infty > p} a_j\Big).$$

Since $n = E(\frac{l-1}{p}) \geq \frac{l}{p} - 2$, we finally obtain :

$$p(l) \leq 2\|X_0\|_h \max\left(a^{-2}, \frac{1}{1-a}\right) \inf_{p \in \mathbb{N}^*} \Big\{a^{l/p} + \sum_{\|j\|_\infty > p} a_j\Big\}. \tag{4.11}$$

An interesting case is obtained when $0 \leq a_i \leq \lambda\,\|i\|_\infty^{-\beta}$ for $i \in \mathbb{Z}^d$ with $\beta > d$ and $0 \leq \lambda < 1$. Then, the proof of Lemma 3 in [13] shows that there exists $C' > 0$ such that

$$p(l) \leq C' \left(\frac{l}{\log l}\right)^{d-\beta}.$$

Consequently, Assumption **(A1)** is satisfied as soon as

$$\beta > d\,\frac{3h-4}{h-2}.$$

4.5 Proof of Theorem 4.2

In order to prove theorem 4.2, we introduce the Berkes and Morrow [4] multiparameter blocking technique. We use the approach and the notations of [3] (see also [6]).

4.5.1 The blocking technique

Let $\alpha > \beta > 1$ be integers to be specified later. Introduce

$$n_0 = 0, \quad n_l = \sum_{i=1}^{l} (i^\alpha + i^\beta), \quad l \in \mathbb{N}^*.$$

Note that $n_l \sim \frac{l^{\alpha+1}}{\alpha+1}$ when $l \to \infty$. For $k \in \mathbb{N}^d$, put $N_k = (n_{k_1}, \ldots, n_{k_d})$. Set

$$B_k = (N_{k-1}, N_k], \quad H_k = \prod_{s=1}^{d} (n_{k_s-1}, n_{k_s-1} + k_s^\alpha], \quad I_k = B_k / H_k.$$

Let $\rho = \tau/8$, L be the set of all indices i corresponding to the "good" blocks $B_i \subset G_\rho$, and H be the set of points in $\mathbb{N}^d$ which belong to some good block. For each point $N = (N_1, \ldots, N_d) \in H$, let $N^{(1)}, \ldots, N^{(d)}$ be defined as follows :

$$N^{(s)}_{s'} = N_{s'},\ s' \neq s, \quad N^{(s)}_s = \min_{n \in H,\, n_{s'} = N_{s'},\, s' \neq s} n_s.$$

We consider also the sets $L_k = \{i : B_i \subset R_k\} \subset \big(L \cap \{i \le k\}\big)$ and $R_k = (M_k, N_k]$ where $M_k = \big((N_k^{(1)})_1, \ldots, (N_k^{(d)})_d\big)$.

Now for $i \in L_k$, let $S^*(H_i) = \sum_{j \in H_i} X_{j,l_i}$ where $l_i = \frac{E\left((i_1 \wedge \ldots \wedge i_d - 1)^\beta\right)}{2}$, where $E(z)$ stands for the integer part of a real number z.
For $N \in G_\tau$ and $N_k < N \le N_{k+1}$,

$$\begin{aligned} S_N &= (S_N - S_{N_k}) + S(R_k) + S((0, N_k]/R_k), \\ W_N &= (W_N - W_{N_k}) + W(R_k) + W((0, N_k]/R_k), \end{aligned}$$

and we use the following decomposition of $S(R_k)$:

$$S(R_k) = \sum_{i \in L_k} (S(H_i) - S^*(H_i)) + \sum_{i \in L_k} S^*(H_i) + \sum_{i \in L_k} S(I_i)$$

The introduction of the vectors $S^*(H_i)$ for $i \in L$ is the key point of our strong approximation since their independence properties will be used to construct an approximation of their partial sums by a sum of Gaussian random vectors.

Lemma 4.1 *$(S^*(H_i))_{i \in L}$ is a family of independent random vectors.*

Proof of lemma 4.1 Let $i, \widetilde{i} \in L$ such that B_i and $B_{\widetilde{i}}$ are two successive blocks, with for example $\widetilde{i}_1 = i_1 + 1$, $\widetilde{i}_s = i_s$ for all $2 \le s \le d$. We have $l_i + l_{\widetilde{i}} \le \frac{E\left((i_1-1)^\beta\right)}{2} + \frac{E\left(i^\beta\right)}{2} \le i_1^\beta$. Since $d(H_i, H_{\widetilde{i}}) \ge i_1^\beta$, we deduce that $H_i^{l_i} \cap H_{\widetilde{i}}^{l_{\widetilde{i}}} = \emptyset$, where for $A \subset \mathbb{Z}^2$ and $\varepsilon > 0$, we use the notation $A^\varepsilon = \{j \in \mathbb{Z}^2 / \, \|j - A\|_\infty < \varepsilon\}$. In the general case, it is obvious that $H_i^{l_i} \cap H_{\widetilde{i}}^{l_{\widetilde{i}}} = \emptyset$ for $i \ne \widetilde{i}$. Since $S^*(H_i)$ is measurable with respect to $\sigma\big(\xi_j, \ j \in H_i^{l_i}\big)$ for all $i \in L$, the result follows from the independence properties of the random field ξ. $\square$

The following lemma provides an important property satisfied by the set G_ρ and will be extensively used in the sequel.

Lemma 4.2 *1) If $M = (M_1, \ldots, M_d) \in G_\rho$, then $M_s \ge [M]^{\frac{\rho}{1+\rho}}$ for $1 \le s \le d$.*
2) There exists $C > 0$ such that for all $i \in L$ and $1 \le s \le d$ then $i_s \ge C[i]^{\frac{\rho}{1+\rho}}$.

In the sequel, we will denote by C a positive constant which can only depend on d, τ, h, $\widetilde{C}$, η and on the covariance function of the field X.

4.5.2 Approximation of $S(R_k)$ by $W(R_k)$

We start by approximating $S(H_i)$ by $S^*(H_i)$.

Lemma 4.3 *Suppose $\beta > \frac{2(1+\rho)}{\rho(2\eta-d)}$. Then there exists $\varepsilon_1 > 0$ such that*

$$\sum_{i\in L_k} \|S(H_i) - S^*(H_i)\| = O\left([N_k]^{1/2-\varepsilon_1}\right).$$

Proof Let q be a real number satisfying $q > -1$ and $q \in \left(\frac{\alpha+1}{2} + \frac{d-2\eta}{2}\frac{\beta\rho}{1+\rho}, \frac{\alpha+1}{2} - 1\right)$. Note that this is possible from the assumptions satisfied by β. For $i \in L_k$, observe that $l_i \geq C[i]^{\frac{\beta\rho}{1+\rho}}$. Then we have from Bienaymé-Tchebichev Inequality and Corollary 4.1 :

$$\begin{aligned}
\mathbb{P}\left(\|S(H_i) - S^*(H_i)\| > [i]^q\right) &\leq [i]^{-2q}\mathbb{E}\,\|S(H_i) - S^*(H_i)\|^2 \\
&\leq [i]^{-2q}\,|H_i| \sum_{j\in\mathbb{Z}^d} \mathbb{E}\left(Y'_{0,l_i} Y_{j,l_i}\right) \\
&\leq C[i]^{-2q+\alpha} l_i^{d-2\eta} \\
&\leq C[i]^{-2q+\alpha+(d-2\eta)\frac{\beta\rho}{1+\rho}}.
\end{aligned}$$

The choice of q implies that $-2q + \alpha + (d-2\eta)\frac{\beta\rho}{1+\rho} < -1$. One can apply the Borel-Cantelli Lemma and we obtain :

$$\|S(H_i) - S^*(H_i)\| \leq C[i]^q \quad \text{a.s.}$$

Hence :

$$\sum_{i\in L_k} \|S(H_i) - S^*(H_i)\| \leq C \sum_{i\in L_k} [i]^q \leq C[k]^{q+1} \leq C[N_k]^{\frac{q+1}{\alpha+1}}.$$

Therefore Lemma 4.3 is satisfied with $\varepsilon_1 = \frac{1}{2} - \frac{q+1}{\alpha+1}$ and $\varepsilon_1 > 0$ from the definition of q. $\square$

Now, we derive the following approximation result :

Lemma 4.4 *If $\alpha - \beta > \frac{2(1+\rho)}{\rho}$ and $\beta > \frac{2(1+\rho)}{\rho(\eta-d)}$, then without changing its distribution, we can redefine the random field X on a rich enough probability space together with a $d-$parameter Wiener processes $W = \left(W_t,\, t \in [0,\infty)^d\right)$ such that there exists $\varepsilon_2 > 0$ satisfying for all $k \in \mathbb{Z}_+^d$ with $L_k \neq \emptyset$,*

$$\sum_{i\in L_k} S^*(H_i) - \Gamma^{1/2} \sum_{i\in L_k} W(B_i) = O\left([N_k]^{1/2-\varepsilon_2}\right).$$

Proof For $i \in L_k$, we define $V_i = \text{Var}\,(S(H_i))$ and $V_i^* = \text{Var}\,(S^*(H_i))$. We first bound the difference $[i]^{-\alpha}V_i^* - \Gamma$. Hence :

$$\left\|[i]^{-\alpha}V_i^* - \Gamma\right\| \leq \left\|[i]^{-\alpha}V_i^* - [i]^{-\alpha}V_i\right\|. + \left\|[i]^{-\alpha}V_i - \Gamma\right\|$$

With Corollary 4.1 :

$$\begin{aligned}\left\|[i]^{-\alpha}V_i^* - [i]^{-\alpha}V_i\right\| &\leq \sum_{j\in\mathbb{Z}^d} \|\Gamma(X_0, X_j) - \Gamma(X_{0,l_i}, X_{j,l_i})\| \\ &\leq C(i_1 \wedge \ldots \wedge i_d)^{\beta(d-\eta)} \\ &\leq C[i]^{\beta(d-\eta)\frac{\rho}{1+\rho}}.\end{aligned}$$

Moreover, from Lemma 4.10 :

$$\left\|[i]^{-\alpha}V_i - \Gamma\right\| = O\left([i]^{-\alpha\frac{\rho}{1+\rho}}\right).$$

This finally leads to :

$$\left\|[i]^{-\alpha}V_i^* - \Gamma\right\| = O\left([i]^{-\frac{\rho}{1+\rho}\min(\alpha,\beta(\eta-d))}\right). \tag{4.12}$$

Now, for $i \in L$, define $S''(H_i) = V_i^{*-1/2}S^*(H_i)$ where $V_i^* = \mathrm{Var}\,(S^*(H_i))$. Note that from (4.12) and assumption **(A3)**, i can be supposed large enough such that $V_i^{*-1/2}$ exists. Define also $S'''(H_i) = \left(V_i^{*1/2} - |B_i|^{1/2}\,\Gamma^{1/2}\right)S''(H_i)$ and $\widetilde{S}(H_i) = |B_i|^{1/2}\,\Gamma^{1/2}S''(H_i)$. Then $S^*(H_i) = S'''(H_i) + \widetilde{S}(H_i)$. Now, $\sum_{i\in L_k} S'''(H_i)$ and $\sum_{i\in L_k}\widetilde{S}(H_i)$ can be bounded as follows :

– Concerning $S'''(H_i)$, let $q > -1$ such that :

$$q < \frac{\alpha-1}{2}, \quad q > \frac{\beta\rho+\alpha}{2(1+\rho)} + \frac{1}{2}, \quad q > \frac{\alpha+1}{2} - \frac{\rho\beta}{2(1+\rho)}(\eta - d).$$

  Note that is possible from assumptions on α and β. Since $\mathbb{E}\left(\|S''(H_i)\|^2\right) = K$, from Bienaymé-Tchebichev Inequality :

$$\begin{aligned}\mathbb{P}\left(|\left\|S'''(H_i)\right\| > [i]^q\right) &\leq [i]^{-2q}\left\|V_i^{*1/2} - |B_i|^{1/2}\,\Gamma^{1/2}\right\|^2 \\ &\leq C[i]^{-2q}\left\|V_i^* - |B_i|\,\Gamma\right\| \\ &\leq C[i]^{-2q}\left(\left\|V_i^* - |H_i|\,\Gamma\right\| + \left|\,|H_i| - |B_i|\,\right|\|\Gamma\|\right).\end{aligned}$$

  We have the following bound :

$$\begin{aligned}\left|\,|H_i| - |B_i|\,\right| &= \left|\prod_{s=1}^{d}(i_s^\alpha + i_s^\beta) - \prod_{s=1}^{d} i_s^\alpha\right| \\ &\leq C\sum_{s'=1}^{d} i_{s'}^\beta \prod_{s\neq s'} i_s^\alpha \\ &\leq C[i]^{\frac{\beta\rho+\alpha}{1+\rho}}.\end{aligned}$$

  We now use (4.12) and we obtain a second bound :

$$\left\|V_i^* - |H_i|\,\Gamma\right\| \leq C[i]^\alpha\left\|\,|H_i|^{-1}\,V_i^* - \Gamma\right\| \leq C[i]^{\alpha-\frac{\rho}{1+\rho}\min(\alpha,\beta(\eta-d))}.$$

Then, assumptions on q imply that there exists $\gamma < -1$ such that :

$$\mathbb{P}\left(\left\|S'''(H_i)\right\| > [i]^q\right) \leq C[i]^\gamma.$$

Therefore we deduce from Borel Cantelli Lemma that :

$$\left\|S'''(H_i)\right\| \leq C[i]^q \quad \text{a.s.}$$

This leads to :

$$\sum_{i\in L_k} \left\|S'''(H_i)\right\| \leq C[k]^{q+1} \leq C[N_k]^{\frac{q+1}{\alpha+1}}. \tag{4.13}$$

Now, define $\varepsilon > 0$ such that $\frac{q+1}{\alpha+1} = \frac{1}{2} - \varepsilon$. Then $\varepsilon > 0$ since $q < \frac{\alpha-1}{2}$.

– We now deal with the term $\sum_{i\in L_k} \widetilde{S}(H_i)$. For this, we use Lemma 4.11, ordering L_k with the lexicographic order for example. Remark that $\mathrm{Var}\left(\widetilde{S}(H_i)\right) = |B_i|\,\Gamma$. Moreover let $p = 2+\delta$, with $\delta > 0$ given in Corollary 4.2. If $a_i = [i]^\gamma$ for all $i \in L_k$, where $\frac{\alpha}{2} + \frac{1}{2+\delta} < \gamma < \frac{1+\alpha}{2}$, then :

$$\begin{aligned}
\sum_{i\in L_k} a_i^{-p}\,\mathbb{E}\left\|\widetilde{S}(H_i)\right\|^p &\leq [i]^{-\gamma(2+\delta)}\left\|\,|B_i|^{1/2}\,\Gamma^{1/2}V_i^{*-1/2}\right\|^{2+\delta}\mathbb{E}\left\|S^*(H_i)\right\|^{2+\delta} \\
&\leq C[i]^{-\gamma(2+\delta)}\,|B_i|^{1+\delta/2}\left\|\,|H_i|^{1/2}\,\Gamma^{1/2}V_i^{*-1/2}\right\|^{2+\delta} \\
&\leq C[i]^{-\gamma(2+\delta)}\,|B_i|^{1+\delta/2} \\
&\leq C[i]^{(\alpha-2\gamma)(1+\delta/2)}.
\end{aligned}$$

The fact that $|H_i|^{1/2}\,\Gamma^{1/2}V_i^{*-1/2}$ is bounded follows from (4.12). By the choice of γ, we have $(\alpha - 2\gamma)(1+\delta/2) < -1$ and the previous sum is finite. Then by Lemma 4.11, we can redefine the sequence $(S'(H_i))_{i\in L}$ on a rich enough probability space together with a sequence $(\zeta_i)_{i\in L}$ of independent $\mathcal{N}(0, |B_i|\,\Gamma)$ $-$random variables such that :

$$\sum_{i\in L_k} \widetilde{S}(H_i) - \sum_{i\in L_k} \zeta_i = O\left([k]^\gamma\right) = O\left([N_k]^{1/2-\varepsilon'}\right), \tag{4.14}$$

where $\varepsilon' = \frac{1}{2} - \frac{\gamma}{1+\alpha}$. Note that $\varepsilon' > 0$ from the definition of γ.

Finally, let $\varepsilon_2 = \varepsilon \wedge \varepsilon'$, and we obtain from (4.13) and (4.14) :

$$\sum_{i\in L_k} \widetilde{S}(H_i) - \sum_{i\in L_k} \zeta_i = O\left([N_k]^{1/2-\varepsilon_2}\right).$$

Now since $\zeta_i = \Gamma^{1/2}\zeta_i'$ for $i \in L$, where $\zeta_i' \sim \mathcal{N}(0, |B_i|\,I_K)$, a straightforward $d-$parameter generalization of Lemma 4 of Csörgo and Révész [8] ensures that we can redefine the sequence $(\zeta_i')_{i\in L}$ on a richer probability space together with a $d-$parameter Wiener process W such that $\zeta_i' = W(B_i)$, for all $i \in L$. The proof is now complete. □

Now to complete the approximation of $S(R_k)$ by $W(R_k)$, we prove the following lemma :

Lemma 4.5 *Suppose that $\alpha - \beta > \frac{2(1+\rho)}{\rho}$, then there exists $\varepsilon_3 > 0$ such that :*

$$\sum_{i \in L_k} S(I_i) = O\left([N_k]^{1/2-\varepsilon_3}\right).$$

Proof of Lemma 4.5 The proof is similar to the proof of Lemma 3.9 of [3]. For $i \in L_k$ we have $I_i = \cup_{s=1}^d I_i(s)$ where $I_i(s)$, $s = 1, \ldots, d$ are disjoint rectangles with

$$|I_i(s)| \leq C i_s^\beta \prod_{s' \neq s} i_{s'}^\alpha.$$

Let $q > -1$ be a real number such that $q \in \left(\frac{\alpha+1}{2} + (\beta - \alpha)\frac{\rho}{1+\rho}, \frac{\alpha-1}{2}\right)$. This is possible from assumptions satisfied by α and β. Denoting $S_i(s) = \sum_{j \in I_i(s)} X_j$, we have :

$$\begin{aligned} \mathbb{P}\left(\|S_i(s)\| \geq [i]^q\right) &\leq C[i]^{-2q} i_s^\beta \prod_{s' \neq s} i_{s'}^\alpha \\ &\leq C[i]^{-2q+\alpha} i_s^{\beta-\alpha} \\ &\leq C[i]^{-2q+\alpha+(\beta-\alpha)\frac{\rho}{1+\rho}}. \end{aligned}$$

Since $-2q + \alpha + (\beta - \alpha)\frac{\rho}{1+\rho} < -1$, the Borel Cantelli lemma and similar calculus as above lead to :

$$\sum_{i \in L_k} \|S_i(s)\| = O\left([N_k]^{\frac{q+1}{\alpha+1}}\right).$$

The choice of $\varepsilon_3 = \frac{1}{2} - \frac{q+1}{\alpha+1} > 0$ leads to the result. □

4.5.3 The remaining terms

We show that the terms $S((0, N_k] \setminus R_k)$, $W((0, N_k] \setminus R_k)$, $S_N - S_{N_k}$, $W_N - W_{N_k}$ are sufficiently small for $N \in G_\tau$. Since all the terms involving the Wiener process are sums of Gaussian i.i.d.r.v. and therefore are $\eta-$weakly dependent random fields, we will only proceed with the partial sums generated by X. We follow [3] and write :

$$\|S((0, N_k] \setminus R_k)\| \leq \sum_{s=1}^d 2^{d-s} D_s(N_k)$$

where $D_s(N_k) = \max_{n \leq N_k^{(s)}} \|S_n\|$. Moreover we have :

$$\max_{N_k < N \leq N_{k+1}} \|S_N - S_{N_k}\| \leq \sum_{\emptyset \neq J \subset \{1,\ldots,d\}} M_k^{(j)},$$

where $M_k^{(J)} = \sup \left\|S(I_k^{(J)})\right\|$ with $I_k^{(J)} = \prod_{s \in J}(n_{k_s}, N_s] \times \prod_{s \notin J}(0, n_{k_s}]$ and the supremum is taken over all N such that $n_{k_s} < N_s \leq n_{k_s+1}$ for all $s \in J$.

Both the following lemmas are now required :

Lemma 4.6 *There exists $\delta > 0$ such that for any $x > 0$ and any block V,*

$$\mathbb{P}\left(M(V) \geq x\sqrt{|V|}\right) \leq Cx^{-2-\delta}.$$

Proof of Lemma 4.6 It follows from Theorem 7.8 and the Markov Inequality. □

Lemma 4.7 *There exists $\gamma_1 > 0$ such that for any block $V = (m, m+n]$ with $n \in G_\rho$:*

$$\mathbb{P}\left(M(V) \geq |V|^{1/2} (\ln |V|)^{d+1}\right) \leq C\, |V|^{-\gamma_1},$$

where $C > 0$ does not depend on m and n.

Proof of Lemma 4.7 The proof is the same as the one of Lemma 7 in [4] using Theorem 7.8 and Proposition 4.3 given in the Annex. □

The two following lemmas are proved analogously to Lemmas 6 and 9 established in [4], using the maximal inequalities given by Lemma 4.6 and Lemma 4.7. Another complete proof with the same notations is given in [7] (see Lemma 2.19 and Lemma 2.20).

Lemma 4.8 *Suppose that $\alpha > \frac{8}{3\tau} - 1$. Then there exists $\varepsilon_4 > 0$ such that for all $N_k \in G_\tau$,*

$$\max_{s=1,\ldots,d} D_s(N_k) = O\left([N_k]^{1/2-\varepsilon_4}\right) \quad a.s$$

Lemma 4.9 *If $\alpha > \frac{2}{\gamma_1}$, where $\gamma_1 \in \mathbb{R}$ is given by Lemma 4.7, then there exists $\varepsilon_5 > 0$ such that for all $N_k \in G_\rho$:*

$$\max_J M_k^{(J)} = O\left([N_k]^{1/2-\varepsilon_5}\right) \quad a.s$$

4.5.4 End of the proof of Theorem 4.2

We choose α and β satisfying :

$$\beta > \frac{2(1+\rho)}{\rho(\eta-d)}, \quad \alpha - \beta > \frac{2(1+\rho)}{\rho}, \quad \alpha > \frac{8}{3\tau} - 1, \quad \alpha > \frac{2}{\gamma_1}.$$

Set $\varepsilon = \min_{i=1,\ldots,5} \varepsilon_i$. Then Theorem 4.2 follows from Lemmas 4.3, 4.4, 4.5, 4.8 and 4.9. □

4.6 Annex

Proposition 4.3 *There exists $\mu_0 > 0$, such that for all $N \in G_\rho$ and $t \in \mathbb{R}^K$,*

$$\left|\mathbb{E}e^{i\,t'\,S_N} - e^{-\|t\|^2/2}\right| \leq C\left(\|t\|^{2+\delta} \vee \|t\|^2\right)[N]^{-\mu_0}.$$

Proof For $i = 1, \dots, d$, let $p_i = E\left(N_i^a\right)$, $q_i = E\left(N_i^b\right)$ and $k_i = E\left(\frac{N_i}{p_i+q_i}\right)$ with $0 < a < b < 1$. For $j \in [0, k-1]$, set :

$$B_j = \prod_{i=1}^{d} ((p_i + q_i)j_i, (p_i + q_i)j_i + p_i].$$

1. With $n = N_1 \wedge \ldots \wedge N_d$ and $l = E(n^b)/2$, we first approximate the random field X by the random field $X_{.,l}$. For $t \in \mathbb{R}^K$, let $f : \mathbb{R}^K \to \mathbb{C}$ such that $f(x) = e^{it'x}$. If for a block U, we write $\widetilde{S}(U) = \sum_{j \in U} X_{j,l}$ and $\widetilde{S}_N = \sum_{j \in (0,N]} X_{j,l}$, then with the point 2 of Corollary 4.1 :

$$\begin{aligned} \left|\mathbb{E}f\left([N]^{-1/2} S_N\right) - \mathbb{E}f\left([N]^{-1/2}\widetilde{S}_N\right)\right| &\leq \|t\|^2 [N]^{-1} \mathbb{E}\left\|S_N - \widetilde{S}_N\right\|^2 \\ &\leq \|t\|^2 \sum_{j \in \mathbb{Z}^d} \left|\mathbb{E}\left(Y'_{0,l} Y_{j,l}\right)\right| \\ &\leq C \|t\|^2 l^{d-2\eta}. \end{aligned}$$

2. We now approximate $\widetilde{S}_N$ by the sum of the independent random vectors $\left(\widetilde{S}(B_j)\right)_{j \in [0,k-1]}$:

$$\begin{aligned} \left|\mathbb{E}f\left([N]^{-1/2}\widetilde{S}_N\right) - \mathbb{E}f\Big([N]^{-1/2} \sum_{j \in [0,k-1]} \widetilde{S}(B_j)\Big)\right| &\leq \|t\|^2 [N]^{-1} \mathbb{E}\Big\|\widetilde{S}_N - \sum_{j \in [0,k-1]} S(B_j)\Big\|^2 \\ &\leq \|t\|^2 [N]^{-1} \left([N] - [k][p]\right) \sum_{j \in \mathbb{Z}^d} \left|\mathbb{E}\left(X'_{l,0} X_{l,j}\right)\right| \\ &\leq C \|t\|^2 n^{(b-a)\vee(a-1)}. \end{aligned}$$

3. We then use the Lindeberg Theorem. Let $(Y_j)_{j \in [0,k]}$ be a finite family of i.i.d. $\mathcal{N}(0, \widetilde{V}_p)$ random variables where $\widetilde{V}_p = \operatorname{Var}(\widetilde{S}(B_j))$. Then, for $\delta \in (0,1)$:

$$\begin{aligned} &\Big|\mathbb{E}f\Big([N]^{-1/2} \sum_{j \in [0,k-1]} \widetilde{S}(B_j)\Big) - \mathbb{E}f\Big([N]^{-1/2} \sum_{j \in [0,k-1]} Y_j\Big)\Big| \\ &\qquad \leq \sum_{j \in [0,k-1]} \mathbb{E}\Big(\big(\frac{\|t\|^3}{2}[N]^{-3/2}\big\|\widetilde{S}(B_j)\big\|^3\big) \wedge \big(2\|t\|^2 [N]^{-1}\big\|\widetilde{S}(B_j)\big\|^2\big)\Big) \\ &\qquad\qquad \leq C \|t\|^{2+\delta} [N]^{-1-\delta/2} \sum_{j \in [0,k-1]} \mathbb{E}\big\|\widetilde{S}(B_j)\big\|^{2+\delta}. \end{aligned}$$

Therefore if δ is small enough, from Proposition 4.2,

$$\Big|\mathbb{E}f\Big([N]^{-1/2} \sum_{j \in [0,k-1]} \widetilde{S}(B_j)\Big) - \mathbb{E}f\Big([N]^{-1/2} \sum_{j \in [0,k-1]} Y_j\Big)\Big| \leq Ct^{2+\delta}[N]^{(a-1)\delta/2}.$$

4. Finally if $V_p = \operatorname{Var}\left(S(B_j)\right)$ and Z is a $\mathcal{N}(0, I_K)$ random variable, then $\sum_{j \in [0,k-1]} Y_j$ has the

same distribution than $\widetilde{V}_p^{1/2}[k]^{1/2}Z$ and

$$\begin{aligned}\left|\mathbb{E}f\left([N]^{-1/2}\widetilde{V}_p^{1/2}[k]^{1/2}Z\right)-\mathbb{E}f\left(\Gamma^{1/2}Z\right)\right| &\leq C\left\|t\right\|^2\left|[N]^{-1}\widetilde{V}_p[k]-\Gamma\right|\\ &\leq C\left\|t\right\|^2[N]^{-1}[k]\left\|\widetilde{V}_p-V_p\right\|\\ &+ C\left\|t\right\|^2\left([p]^{-1}V_p\frac{[N]-[k][p]}{[N]}+\left\|[p]^{-1}V_p-\Gamma\right\|\right).\end{aligned}$$

Now,

– Using Corollary 4.1, we have

$$[N]^{-1}[k]\left\|\widetilde{V}_p-V_p\right\|\leq\frac{[k][p]}{[N]}\sum_{j\in\mathbb{Z}^d}\left\|\Gamma\left(X_0,X_j\right)-\Gamma\left(X_{l,0},X_{l,j}\right)\right\|=O\left(l^{d-\eta}\right).$$

–

$$[p]^{-1}\left\|V_p\right\|\frac{[N]-[k][p]}{[N]}\leq\sum_{j\in\mathbb{Z}^d}\left\|\Gamma\left(X_0,X_j\right)\right\|\frac{[N]-[k][p]}{[N]}=O\left(n^{(b-a)\vee(a-1)}\right).$$

– As a consequence of Lemma 4.10, we have

$$\left\|[p]^{-1}V_p-\Gamma\right\|=O\left(n^{(-a)\vee a(d-\eta)}\right)=O\left(n^{-a}\right).$$

We have finally shown that if $\mu=(b-a)\vee(a-1)\frac{\delta}{2}\vee b(d-\eta)$:

$$\left|\mathbb{E}f\left([N]^{-1/2}S_N\right)-\mathbb{E}f\left(\sigma Z\right)\right|\leq C\left\|t\right\|^{2+\delta}\vee\left\|t\right\|^2n^{\mu}.$$

Since $N\in G_\rho\Longrightarrow n\geq[N]^{\frac{\rho}{1+\rho}}$, the result follows by setting $\mu_0=-\frac{\rho}{1+\rho}\mu$. □

Lemma 4.10 *Under assumption* **(A1)**, *then for all* $N=(N_1,\ldots,N_d)\in(\mathbb{N}^*)^d$

$$\left\|[N]^{-1}\,Var\,(S_N)-\Gamma\right\|=O\left(l^{-1}\right),$$

where $l=\min_{s=1,\ldots,d}N_s$.

Proof of lemma 4.10

– First,

$$\Gamma=\sum_{|i_1|<N_1,\ldots,|i_d|<N_d}\Gamma(X_0,X_i)+A,$$

and using the point 1 of Proposition 4.1, we have for a suitable constant $K>0$:

$$\begin{aligned}\|A\| &\leq \sum_{\|i\|_\infty>l}\Gamma(X_0,X_i)\\ &\leq C\sum_{k\geq l}k^{d-\eta-1}.\end{aligned}$$

Therefore $A=O\left(l^{d-\eta}\right)$.

– Moreover,

$$\mathrm{Var}\,(S_N) = \sum_{|i_1|<N_1,\ldots,|i_d|<N_d} (N_1 - |i_1|)\cdots(N_d - |i_d|)\,\Gamma(X_0, X_i),$$

and one can deduce that :

$$\left\| [N]^{-1}\mathrm{Var}\,(S_N) - \sum_{|i_1|<N_1,\ldots,|i_d|<N_d} \Gamma(X_0, X_i) \right\| \leq (2^d - 1)l^{-1} \sum_{i\in\mathbb{Z}^d} \|\Gamma(X_0, X_i)\|\,.$$

under assumption **(A1)** $\eta - d > 1$, and hence the result of the lemma follows from the two last points.
□

Lemma 4.11 *(see [15]) Let $(Z_l)_{l\in\mathbb{N}^*}$ be a sequence of independent $\mathbb{R}^k$ valued random vectors with zero means and $Cov(Z_l) = \sigma_l^2\Gamma$ for $l \in \mathbb{N}^*$. Assume that there exist $q \in (2, 4)$ and a sequence $(a_l)_{l\in\mathbb{N}^*}$ such that $a_l \to \infty$ $(l \to \infty)$, satisfying*

$$\sum_{l=1}^{\infty} \frac{\mathbb{E}\,\|Z_l\|^q}{a_l^q} < \infty.$$

Then on a richer probability space we can construct a sequence of independent normal random vectors $(\zeta_l)_{l\in\mathbb{N}^*}$ with $\mathbb{E}\zeta_l = 0$ and $Cov(\zeta_l) = \sigma_l^2\Gamma$ for $l \in \mathbb{N}^*$, such that

$$\sum_{l=1}^{n} Z_l - \sum_{l=1}^{n} \zeta_l = o\,(a_n) \qquad a.s.$$

Bibliographie

[1] Andrews, D. (1984) *Non strong mixing autoregressive processes.* Journal of Applied Probability 21, 930-934.

[2] Aue, A., Berkes, I., Horváth, L. (2006) *Strong approximation for sums of squares of augmented GARCH sequences.* Bernoulli, 12, pp. 583-608.

[3] Balan, R. (2005) *A strong invariance principle for associated random fields.* The Annals of Probability, Vol. 33, No. 2, 823-840.

[4] Berkes, I., Morrow, G. J. (1981) *Strong invariance principle for mixing random fields.* Z. Wahrsh. Verw. Gebiete 57 15-37.

[5] Berkes, I., Philipp, W. (1979) *Approximation theorems for independent and weakly dependent random vectors.* The Annals of Probability, Vol.7, No. 1, 29-54.

[6] Bulinski, A., Shashkin, A. (2006)*Strong invariance principle for dependent random fields* IMS Lecture Notes–Monograph Series, Vol. 48, 128-143.

[7] Bulinski, A., Shashkin, A. *Limit Theorems for Associated Random Fields and Related Systems.* Advanced Series on Statistical Science and Applied Probability. Vol 10.

[8] Csörgo, M., Révész, P. (1975) *A new method to prove Strassen type laws of invariance principle.* I. Z. Wahrsch. Vervw. Gebiete 31 255-260.

[9] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, and S., Prieur, C. (2007) *Weak dependence, examples and applications*, Lecture Notes in Statistics 190.

[10] Doukhan, P. and Lang, G. (2002) Rates in the empirical central limit theorem for stationary weakly dependent random fields, Statistical Inference Stochastic Processes, Vol. 5, 199-228.

[11] Doukhan, P., Louhichi, S. (1999) *A new weak dependence condition and applications to moment inequalities.* Stoch. Proc. Appl. 84, 313-342.

[12] Doukhan, P., Mayo, N., Truquet, L. (2008) *Weak dependence, models and some applications.* To appear in Metrika.

[13] Doukhan, P., Truquet, L. *A fixed point approach to model random fields.* Alea 3, 111-132.

[14] Doukhan, P. and Wintenberger, O. (2007) A central limit theorem under non causal weak dependence and sharp moment assumptions, Prob. Math. Stat. 27, 45-73.

[15] Einmahl, U. *A useful estimate in the multidimensional invariance principle.* Probab. Theory Relat. Fields. 76 : 81-101

[16] Guyon, X. (1995) *Random fields on a network : modeling, statistics, and applications.* Graduate Texts in Mathematics. Springer-Verlag, New-York.

[17] Illig, A., Truong-Van, B. (2006)*Asymptotic results for spatial ARMA models.* Commun. Stat., Theory Methods 35, No.4, 671-688.

[18] Liu, W., Lin, Z. (2008) *Strong approximation for stationary processes.*

[19] Móricz, F. (1983) *A general moment inequality for the maximum of the rectangular partial sums of multiple series.* Acta Math. Hung., 41, 3-4, pp. 337-346.

[20] Wu, W.B. (2007) *Strong invariance principles for dependent random variables.* The Annals of Probability, Vol 35, nř6, pp. 2294-2320.

Chapitre 5

A nonparametric resampling for non causal random fields and its application to the texture synthesis

Abstract

We study an extension to non causal Markov random fields of the resampling scheme given in Bickel et Levina (2006)[5] for texture synthesis with Markov mesh models. This extension is similar to a nonparametric method proposed by Paget and Longstaff (1998)[19] for texture synthesis and we also use their multiscale synthesis algorithm incorporating local annealing. We discuss some statistical properties and theoretical points for the convergence of the procedure and provide several convincing simulation examples.

5.1 Introduction

Over the two last decades, there was a particular attention to study the problem of texture synthesis. The goal of texture synthesis can be stated as follows : Given a texture sample, synthesize a new texture that appears similar to a human observer. Texture mapping or image compression are frequent applications for such algorithms. The stochastic nature of texture variations makes it a particularly natural area for applying statistical methods. The pioneer work of Cross and Jain (1983) [6] have shown the ability of Markov random fields to model a homogeneous texture. Such parametric models have been used for texture synthesis (as in [21]), but they require the estimation of a high number of parameters for capturing the complexity of real textures, which leads to computational difficulties. On the other hand, some algorithms model textures as a set of features, and generate new images by matching the features in an example texture ([7], [15], [23]). Those methods work very well for

stochastic textures but have sometimes difficulties with highly structured ones.
Significant advances have been done in the area of texture synthesis using nonparametric algorithms with Markov random fields. A popular algorithm has been introduced by Efros and Leung (1999) [10]. Many variations of their method have been published that speed up and optimize the original algorithm in different ways in Wei et Levoy (2000) [24], Efros and Freeman (2001) [9] and Liang et al. (2001)[16] among others. The main statistical idea behind those algorithms is to consider the observed texture as a realization of a strictly stationary MRF. The data are used to construct an estimate of a local conditional distribution function of the field and a new texture is synthetized with a simulation procedure. Typically, the synthesis starts using a seed and pixels are synthetized in a given order by a recursive simulation of the random field as for time series. Intrinsically, those simulation procedures suppose the causality nature of the observed stochastic process. At a first sight this dependence form seems unnatural but the above algorithms work well on a wide variety of textures which seem well approximated by such random fields.
On the other hand, some noncausal procedures have been investigated. The FRAME model introduced by Zhu, Wu and Mumford [25] combine noncausal MRF models and feature matching. This last model has a mathematical justification : maximum entropy with empirical histograms of a finite number of filter responses are used to derive a parametric MRF for the whole distribution of the texture. Despite its solid statistical modeling, FRAME models does not work always very well on real textures.
Paget and Longstaff (1998)[19] have considered another algorithm, with a nonparametric noncausal MRF. Contrarily to [25], the random field is specified through the conditional distribution and the empirical histogram is smoothed with a kernel which allows a simulation procedure with the Gibbs sampler. To avoid long relaxation time and phase discontinuities, Paget and Longstaff have used multiscale grids and have incorporated a temperature parameter for the pixels and the resulting algorithm is shown to be able to synthetize stochastic textures but also highly structured ones.
Except the work of Zhu et al.[25], not many theoretical works have been developed to study the consistency of such procedures. To our knowledge, the only contribution is the work of Bickel et Levina [5] who define a formal bootstrap scheme for resampling stationary (causal) random fields which gives a theoretical justification to the algorithm of Efros and Leung [10].
The goal of this paper is to extend the method of Bickel and Levina to noncausal random fields for modeling textures as in [19]. Of course, the use of the Gibbs sampler leads to long computational times and this gives a clear advantage to causal algorithms. However, from a theoretical point of view, the class of noncausal Markov random fields is known to be wider than the class of causal fields and in fact only a noncausal field has a real physical sense. In [5], the authors study a nonparametric estimation of the local conditional distribution function associated to the random field which is used to simulate an approximate causal field. This method is an extension to random fields of a $p-$order

Markov bootstrap algorithm for time series [20]. We will use the same nonparametric estimation of a conditional law and we will use the multiscale synthesis algorithm given in [19] to give simulation examples.
The paper is organized as follows. In the following Section 5.2, we recall the results of Bickel et Levina and provide the natural extension of their method to the noncausal case. Some considerations on the convergence and the convergence rate of such algorithm are also provided. Section 5.3 is devoted to recall the multiscale algorithm used by Paget and Longstaff and we incorporate our bootstrap method to provide several simulation examples. Theoretical investigations are postponed to the two last sections of the paper.

5.2 The Markov Mesh Models algorithm

5.2.1 Principle

We first recall the Markov Mesh Models (MMM in sequel) algorithm introduced by Bickel and Levina [5]. This algorithm is different of the original algorithm of Efros and Leung [10] by the order in which pixels are filled in the synthesized texture (raster instead of spiral), and the weights with which the pixels are resampled. One can note that the raster order is used in some variations of the original algorithm (see [24] and [16]).

In all the sequel we consider $\{X_t,\, t\in\mathbb{N}^*\times\mathbb{N}^*\}$ a real-valued random field and a positive integer $o\in\mathbb{N}^*$. We will use the following notations :
– for $A\subset\mathbb{N}^*\times\mathbb{N}^*$, X_A denote the family $(X_t)_{t\in A}$;
– for $A,\, B\subset\mathbb{N}^*\times\mathbb{N}^*$, $A+B=\{t_A+t_B,\,(t_A,t_B)\in A\times B\}$ and $A-B=\{t_A-t_B,\,(t_A,t_B)\in A\times B\}$.
For $t=(t_1,t_2)\in\mathbb{N}^*\times\mathbb{N}^*$ and $s\in\mathbb{N}^*\times\mathbb{N}^*$, define the index sets

- $U_t^{(o)}=\{u=(u_1,u_2)\in\mathbb{N}^*\times\mathbb{N}^*;\ \max(1,t_1-o)\le u_1\le t_1,\,\max(1,t_2-o)\le u_2\le t_2 \text{ and } u\ne t\}$;
- $U_t^{(o)}(s)=U_t^{(o)}-\{t\}+\{s\}$;
- $W_t=\{1,\dots,t_1\}\times\{1,\dots,t_2\}\setminus\{t\}$.

The set $U_t^{(o)}$ is always included in the square of size $(o+1)\times(o+1)$ with t as the bottom right corner, t itself excluded, but there are $(o+1)^2-1$ possible shapes of $U_t^{(o)}$. Then,

Definition 5.1 *A random field $X=\{X_t,\, t\in\mathbb{N}^*\times\mathbb{N}^*\}$ is a Markov mesh model if there exists $o\in\mathbb{N}^*$ such that for all $t\in\mathbb{N}^*\times\mathbb{N}^*$,*

$$\mathbb{P}\left(X_t/X_{W_t}\right)=\mathbb{P}\left(X_t/X_{U_t^{(o)}}\right). \tag{5.1}$$

Now, the MMM resampling algorithm of Bickel and Levina [5] can be presented. First assume that a trajectory of a MMM X is observed on the index set $\{1,\dots,T_1\}\times\{1\dots,T_2\}$ with $T_1,T_2\in\{o,o+1,\dots\}\times\{o,o+1,\dots\}$, *i.e.*

$$(X_t,\, t\in\{1,\dots,T_1\}\times\{1\dots,T_2\})$$

is known. Then consider a family of kernels $(W^{(\ell)})_{\ell\in\mathbb{N}^*}$ that are Borelian functions $W^{(\ell)}:\mathbb{R}^\ell\to[0,\infty)$ satisfying some general smoothness assumptions (see Assumption **(A4)** below). Moreover, for a resampling width $b>0$ and all $\ell\in\mathbb{N}^*$, define

$$W_b^{(\ell)}(y)=b^{-\ell}W^{(\ell)}(y/b)\quad\text{for all } y\in\mathbb{R}^\ell.$$

In the sequel, for simplicity, we will omit the exponent o and ℓ for respectively $U_t^{(o)}$, $U_t^{(o)}(s)$ and $W_b^{(\ell)}$.

The MMM resampling algorithm

In the sequel we will denote $X^*=\{X_t^*,\, t\in\mathbb{N}^*\times\mathbb{N}^*\}$ the generated texture from $(X_t,\, t\in\{1,\dots,T_1\}\times\{1\dots,T_2\})$. There are 3 main steps in this algorithm :

1. Select a starting value for $\{X_t^*: 1\le t_1\le o+1, 1\le t_2\le o+1\}$, the top left $(o+1)\times(o+1)$ square. Typically the starting value will be a $(o+1)\times(o+1)$ square random chosen from the observed field $(X_t,\, t\in\{1,\dots,T_1\}\times\{1\dots,T_2\})$.
2. Suppose that there exists $(u,v)\in\mathbb{N}^*\times\mathbb{N}^*$ such that X_t^* has been generated for $t\in\{1,\dots,u-1\}\times\{1,\dots,v\}\cup\{u\}\times\{1,\dots,v-1\}$, that is, $u-1$ rows are filled in completely, and the row u is filled up for the column v. To generate the next value $X_t^*=X_{(u,v)}^*$, let N_t be a discrete random variable with probability distribution

$$\mathbb{P}(N_t=s)=\frac{1}{Z}W_b\big(X_{U_t}^*-X_{U_t(s)}\big)$$

   for all $s\in\mathbb{N}^*\times\mathbb{N}^*$ such that $U_t(s)\subset\{1,\dots,T_1\}\times\{1\dots,T_2\}$ and where $Z=\sum_s W_b(Y_t^*-Y_t(s))$ is a normalizing constant. Note that the set of all possible s is such that all locations where the conditioning neighborhood fits within the observed texture field.
3. Generate N_t and set $X_t^*=X_{(u,v)}^*=X_{N_t}$.

In Figure 7.2, we show two steps in the progress of the MMM algorithm, just after the choice of the seed (step 1 above) and when the neighborhood of the pixel is full (see the center of the picture in Figure 7.2). In Figure 5.2, we illustrate the difference with the original algorithm of Efros and Leung. Here, the seed is put in the center of the new texture and the synthesis is done with a spiral ordering. To synthesize a pixel at a site t, one considers only the value of pixels already synthesized in a given square window centered at t. Another difference with the MMM algorithm is the choice

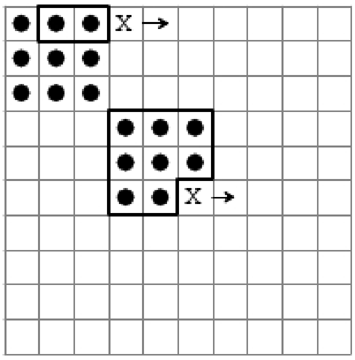


FIG. 5.1: MMM algorithm, $o = 2$

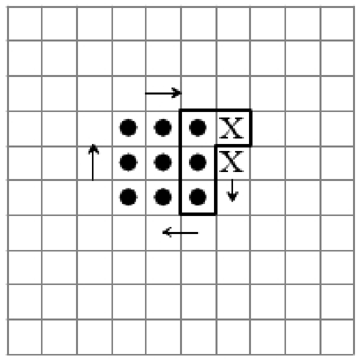


FIG. 5.2: Efros and Leung algorithm, $o = 2$

of uniform weights for the synthesis (see [5] for details). The MMM algorithm is formulated for a particular class of random fields, the Markov Mesh Models (also known as Picard random fields) which were introduced by Abend, Harley and Kanal [1]. These models have been developed for image applications and can be simulated recursively an quickly. The resampling scheme described above is an adaption of a method proposed for bootstrapping Markovian time series ([22], [20]).

5.2.2 Consistency results for causal models

First, let us introduce new notations :
- for $A \subset \mathbb{Z}^2$, $|A|$ is the cardinal of A.
- for $x = (x_1, x_2) \in \mathbb{Z}^2$, $\|x\|_\infty = \max(|x_1|, |x_2|)$.
- for $A, B \subset \mathbb{Z}^2$, $d(A, B) = \inf_{x \in A, y \in B}\{\|x - y\|_\infty\}$.
- for $y \in \mathbb{R}^\ell$ with $\ell \in \mathbb{N}^*$, $\|y\|$ is the usual Euclidian norm.
- for $T = (T_1, T_2) \in \mathbb{N}^* \times \mathbb{N}^*$, let $[T] = T_1 T_2$ and $T \to \infty$ means $T_1 \wedge T_2 \to \infty$.

Let $A \in \mathbb{Z}^2$ such that $|A| < \infty$. For $t \in \mathbb{Z}^2$, define

$$Y_t = (X_{t+j})_{j \in A}\,.$$

Moreover we define the following subsets of $\mathbb{N}^* \times \mathbb{N}^*$:

$$I_T = \Big\{t \in \{1, \dots, T_1\} \times \{1 \dots, T_2\},\ \{t\} - A \subset \{1, \dots, T_1\} \times \{1 \dots, T_2\}\Big\}.$$

To show the consistency of their algorithm, Bickel et Levina have proved a general lemma about the estimation of the local conditional distribution function

$$F_{X/Y}(x/y) = \mathbb{P}(X_t \leq x/Y_t = y)$$

(see Theorem 2 in [5]). We first recall this theorem and its assumptions.

Assumptions of Theorem 2 in [5]

(A1) The random field X is strictly stationary and α-mixing, *i.e.* if for $k, u, v \in \mathbb{N}^*$,

$$\alpha_X(k,u,v) = \sup_{E,F \in \mathbb{Z}^2, d(E,F)=k, |E|=u, |F|=v} \left\{ |P(AB) - P(A)P(B)|,\ A \in \sigma(X_E), B \in \sigma(X_F) \right\}$$

are the strong mixing coefficients such that there exist $\varepsilon > 0$, $\tau > 2$ satisfying for all integers $u, v \geq 2$, $u + v \leq c$, where c is the smallest even integer such that $c \geq \tau$,

$$\sum_{k=1}^{\infty} (k+1)^{2(c-u+1)-1} \alpha_X(k,u,v)^{\varepsilon/(c+\varepsilon)} < \infty.$$

(A2) X_t has a compact support $S \subset \mathbb{R}$.

(A3) $F_{X,Y} = \mathbb{P}(X_t \leq \cdot, Y_t \leq \cdot)$, $F_{X/Y}$ and $F_Y = \mathbb{P}(Y_t \leq \cdot)$ have bounded continuous strictly positive densities (denoted $f_{X,Y}$, $f_{X/Y}$ and f_Y respectively) with respect to Lebesgue measure. Moreover, there exists $L > 0$ such that for any $y, y' \in S^A$, any $x \in S$,

$$\left| \int_{-\infty}^{x} f_{X,Y}(z,y)dz - \int_{-\infty}^{x} f_{X,Y}(z,y')dz \right| \leq L \left\| y - y' \right\|.$$

(A4) The family of kernels $(W^{(\ell)})_{\ell \in \mathbb{N}^*}$ is such that $W^{(\ell)} : \mathbb{R}^\ell \to (0, \infty)$ are bounded, symmetric and first-order Lipshitz continuous functions such that for all $\ell \in \mathbb{N}^*$,

$$\int u\, W^{(\ell)}(u) d\lambda_\ell(u) = 0 \quad \text{and} \quad \int \|u\|\, W^{(\ell)}(u) d\lambda_\ell(u) < \infty.$$

Moreover, the width of $W_b^{(\ell)}$ is supposed to be such that $b = b_T = O([T]^{-\delta})$, with $\delta > 0$.

To show the consistency of the MMM resampling algorithm, Bickel and Levina have established the convergence of the following sample cumulative conditional distribution function, that is, for $(x,y) \in S \times S^A$ and $T \in \mathbb{N}^* \times \mathbb{N}^*$ such that $I_T \neq \emptyset$:

$$F_T(x/y) = \frac{1}{Z_T} \sum_{s \in I_T} \mathbb{1}_{X_s \leq x} W_{b_T}(y - Y_s), \tag{5.2}$$

where $Z_T = \sum_{s \in I_T} W_{b_T}(y - Y_s)$.

Theorem 5.1 (Theorem 2 [5]) *If X is a MMM satisfying assumptions $(A1)-(A4)$, then for all $A \in \mathbb{Z}^2$ such that $|A| < \infty$,*

$$\sup_{(x,y)\in S\times S^A} \left|F_T(x/y) - F_{X/Y}(x/y)\right| \to_{T\to\infty} 0.$$

Theorem 5.1 shows the uniform convergence of the conditional distribution of a pixel given its neighborhood. Using this general result with neighborhoods A of causal nature (e.g. $A = U_t - \{t\}$), Bickel and Levina show the consistency of their MMM algorithm and also of the original spiral resampling algorithm of Efros and Leung. Their proof use the conditional independence properties of the MMM which allow a recursive computation of the joint laws (we refer to Theorem 1 in [5] for details). This resampling scheme uses the kernel regression estimation and requires some regularity assumptions (see Assumptions **(A1-4)**). As it is pointed in [5], those assumptions are perfectly plausible for most real textures : the mixing property is natural for stochastic textures, the compactness assumption is always satisfied since the number of gray levels is finite, and this number is sufficiently high in most of real textures to make the smoothness assumptions plausible.
However causal MMM are not really an appropriated for modeling texture : indeed, Why to choose a certain direction as for the dependence of the field ? It is more natural to consider a spatial model for which there are no privileged direction for the dependence, *i.e.* a noncausal random field.

5.2.3 An extension to the noncausal case and a convergence rate of Theorem 5.1

To extend the previous results of [5] to noncausal fields, consider the following neighborhood $\mathcal{N}_o$ where $o \in \mathbb{N}^*$

$$\mathcal{N}_o = \{j \in \mathbb{Z}^2 / 0 < \|j\|_\infty \leq o\}.$$

Thus $\{t\} + \mathcal{N}_o$ is the natural extension of the set $U_t^{(o)}$ in the noncausal case. Denote

$$v = |\mathcal{N}_o| = (2o+1)^2 - 1.$$

The MMM is a very particular case of Markov random fields. If $X = \{X_t,\, t \in \mathbb{Z}^2\}$ is a $\mathbb{R}$-valued random field, then :

Definition 5.2 *$X = \{X_t,\, t \in \mathbb{Z}^2\}$ is a Markov random field if there exists $o \in \mathbb{N}^*$ such that for all $t \in \mathbb{Z}^2$,*

$$\mathbb{P}\left(X_t / X_{\mathbb{Z}^2\setminus\{t\}}\right) = \mathbb{P}\left(X_t / X_{t+\mathcal{N}_o}\right). \tag{5.3}$$

We will again assume that $(X_t,\, t \in \{1,\dots,T_1\} \times \{1\dots,T_2\})$ is known. Then, for all $t \in \mathbb{Z}^2$, define now :

$$Y_t = (X_{t+j})_{j\in\mathcal{N}_o} = X_{t+\mathcal{N}_o}.$$

First, a convergence rate for the Theorem 2 of [5] can be established and it is also satisfied in the noncausal case :

Theorem 5.2 *If X is a is a noncausal Markov random field satisfying assumptions* $(\mathbf{A1-4})$*, then for all $A \in \mathbb{Z}^2$ such that $|A| < \infty$,*

$$\sup_{(x,y)\in S\times S^{\mathcal{N}_o}} \left|F_T(x/y) - F_{X/Y}(x/y)\right| = O\left([T]^{-\gamma}\right) \quad a.s$$

where $0 < \gamma < \dfrac{\tau-2}{2(v+1)(\tau+v+2)}$ *and* $b = b_T = O([T]^{-\delta})$ *with* $\delta = \dfrac{\tau-2}{2(v+1)(\tau+v+2)}$.

Since MMM is a particular case of Markov random field, this result is also satisfied by MMM. It is interesting to see in both the causal or noncausal cases that the convergence rate of the MMM resampling algorithm is depending on a power law of $[T]$ (even if the choice of the optimal bandwidth b_T is depending on unknown parameters τ and v). Moreover the maximal exponent of convergence rate that we can obtain in Theorem 5.2 is $\frac{1}{2(1+v)}$ (that requires $\tau \to \infty$ for the mixing assumption). Remark that if $o = 0$ (corresponding to a independent random field) then $v = 0$ and the convergence rate is arbitrary close to $T^{1/2}$.

A partial consistency result for the MMM resampling algorithm in the noncausal case

In order to extend to the noncausal case the consistency proof of Bickel et Levina for the MMM resampling algorithm, we define the following one point conditional distribution defined by :

$$F_T(dx/y) = \frac{1}{Z_T}\sum_{s\in I_T} W_{b_T}\left(y - Y_s\right)\delta_{X_s}(dx), \tag{5.4}$$

where δ_x is the usual Dirac mass measure. Note that (5.4) is equal to (6.5) in the case $A = \mathcal{N}_o$. However and contrary to the causal case, the one point distribution (5.4) cannot be in general the one point conditional distribution of a noncausal Markov random field (nevertheless, we will use (5.4) to run a Gibbs sampler). A statistical problem with texture modeling by a noncausal Markov random field is to define a consistent nonparametric estimate of the one point conditional distributions which is also compatible with the existence of a conditional specification. This would allow to define an approximate Markov random field. We did not found a such estimate. Some tools are given in the Annex about the link between the convergence of a sequence of one point conditional distributions and the behavior of their joint laws provided they are well defined (see Theorem 5.3 and Theorem 5.4). Here we only provide a restrictive result of consistency of a simulation procedure directly with the Markov chain linked to the Gibbs sampler.

Suppose that we use the conditional distributions (5.4) and the Gibbs sampler to synthetize a new texture on a rectangle $R = R_T = \{1,\ldots,u_T\} \times \{1,\ldots,v_T\}$. We suppose here that assumptions of Theorem 5.1 hold. Though the conditional distributions F_T defined in (5.4) are not compatible with a Markov random field, we can use those distributions to simulate a Markov chain. We denote by $\prec$ the lexicographic order relation on R. Let z is an arbitrary element of $S^{\mathbb{Z}^2}$ not depending on T. If $x, y \in S^R$ and $s \in R$, we define the vectors $yx(s) \in S^{\mathcal{N}_o}$ such that $yx(s)_j = y_{s+j}$ if $s+j \prec s$ and $yx(s) = x_{s+j}$ otherwise, completed with the boundary conditions $x_{s+j} = z_{s+j}$ or $y_{s+j} = z_{s+j}$ if site $(s,j) \in R \times \mathcal{N}_o$ is such that $s+j \notin R$.

Now for $T \in \mathbb{N}^* \times \mathbb{N}^*$ such that $T_1, T_2 \geq 2o+1$ (this ensures that I_T is not empty), we define the following transition on $X^R_{I_T} \subset S^R$:

$$P_T(x, dy) = \otimes_{s \in R} F_T\left(dy_s / yx(s)\right).$$

Note that P_T corresponds to the transition of the homogeneous Markov chain associated to the conditional distributions F_T when we implement the Gibbs sampler with a raster ordering for the visiting scheme (see [14] Theorem 6.2.1). Now as for the classical Gibbs sampler, we simulate a Markov chain on I_T^R, with initial value $w \in I_T^R$ and transition P_T. Since P_T is a positive transition, the law of this Markov chain with finite state space converges to its unique invariant probability denoted by μ_T. Then we have the following equality :

$$\mu_T(A) = \int P_T(x, A)\mu_T(dx), \qquad A \in \mathcal{B}(S^R),$$

where $\mathcal{B}(S^R)$ denote the Borel $\sigma-$algebra on S^R. One can mention that μ_T is not in general a measure that admits F_T as conditional distributions.
Since S^R is a compact metric space, the tightness of the sequence $(\mu_T)_T$ implies the existence of a cluster point denoted by μ. We are going to show that $\mu = \mu_R$, where μ_R denotes the conditional law $X_R / X_{\partial R} = z_{\partial R}$, where $\partial R = (R + \mathcal{N}_o) \setminus R$. Then by uniqueness of the cluster point, we will deduce the following consistency result :

$$\text{Almost surely : } \lim_{T \to \infty} \mu_T = \mu_R \quad \text{in distribution.} \tag{5.5}$$

To show that $\mu = \mu_R$, we first observe that μ_R is an invariant probability of the transition P on S^R defined by

$$P(x, dy) = \otimes_{s \in R} F_{X/Y}\left(dy_s / yx(s)\right), \quad x \in S^R.$$

Then P define a positive Markov chain and μ_R is the unique invariant probability. In fact P is the transition of the homogeneous Markov chain defined in the Gibbs sampler for the simulation of a realization of μ_R (still in the case of a periodic visiting scheme). Then if we prove that $\mu(A) = \int P(x, A)\mu(dx)$, $\forall A \in \mathcal{B}\left(S^R\right)$, we we can conclude that $\mu = \mu_R$.

Suppose that $(T_n)_{n\in\mathbb{N}}$ is sequence in $\mathbb{N}^*\times\mathbb{N}^*$ such that $\lim_{n\to\infty}\mu_{T_n}=\mu$. Then if g be a continuous and bounded function on S^R. We have :

$$\begin{aligned}\left|\int g(y)\mu_{T_n}(dy)-\int g(y)P(x,dy)\mu(dx)\right| &= \left|\int g(y)P_{T_n}(x,dy)\mu_{T_n}(dx)-\int g(y)P(x,dy)\mu(dx)\right|\\ &\le \sup_{x\in S^R}\left|\int g(y)P_{T_n}(x,dy)-\int g(y)P(x,dy)\right|\\ &+ \left|\int g(y)P(x,dy)(\mu_{T_n}-\mu)(dx)\right|\\ &= A+B.\end{aligned}$$

Since the function $x\to\int g(y)P(x,dy)$ is still bounded and continuous from assumption **(A2)**, the weak convergence of the sequence $(\mu_{T_n})_n$ implies that $B\to 0$ $(n\to\infty)$.
Now we show that $A\to 0$ $(n\to\infty)$. First we observe that if h is a continuous and bounded function on $S^R\times S^R$ and $s\in R$, then :

$$\sup_{(x,y)\in S^R\times S^R}\left|\int h(x,y)\left(F_{T_n}(dy_s/yx(s))-F_{X/Y}(dy_s/yx(s))\right)\right|\to_{n\to\infty}0,\quad \text{a.s.}\tag{5.6}$$

The proof of (5.6) is omitted since the proof is very similar to the assertion $A\to 0$ in the proof of Theorem 5.3, using Theorem 5.1. If $u\in R$ is such that $u\succ s$, $\forall s\in R\setminus\{u\}$, we have :

$$\begin{aligned}&\left|\int g(y)P_{T_n}(x,dy)-\int g(y)P(x,dy)\right|\\ &\le\left|\int\int g(y)F_{X/Y}(dy_u/yx(u))\left(\otimes_{s\in R\setminus\{u\}}F_{T_n}(dy_s/yx(s))-\otimes_{s\in R\setminus\{u\}}F_{X/Y}(dy_s/yx(s))\right)\right|\\ &\qquad+\sup_{y_s\in S,s\neq u}\left|\int g(y)\left(F_{T_n}(dy_u/yx(u))-F_{X/Y}(dy_u/yx(u))\right)\right|.\end{aligned}\tag{5.7}$$

Then if we iterate the bound (5.7), using a non increasing enumeration of the sites of R, the convergence $A\to 0$ follows from a repeated use of (5.6). Then, by the uniqueness of the limit of the sequence $(\mu_{T_n})_n$, we conclude that $\mu(dy)=\int P(x,dy)\mu(dx)$ and the convergence (5.5) follows from the previous remarks.

However, for obtaining the consistency the natural asymptotic requires that R_T increases to $\mathbb{Z}^2$. Unfortunately, we did not find a proof in this case.

5.3 The approach of Paget and Longstaff and simulation examples

5.3.1 Paget and Longstaff method

In their paper, Paget and Longstaff [19] have proposed a noncausal estimate of the local conditional distribution similar than ours, using also a kernel which smooths the multidimensional histogram.

Our approach is more linked to the idea of a resampling scheme and appears in a natural way from the nonparametric estimation of the conditional expectation $(x, y) \mapsto \mathbb{P}\left(X_t \leq x / Y_t = y\right)$. The Gibbs sampler (see [12]) is a classical stochastic relaxation (SR) algorithm which is used for the simulation of Markov random fields. But as pointed in [19], a problem with the single-scale relaxation process is that global image characteristics evolve indirectly in the relaxation process. Global image characteristics are typically only propagated across the image lattice by local interactions and therefore evolve slowly, requiring long relaxation times to obtain equilibrium. Moreover the conditional distribution given in (5.4) requires the comparaison of a neighborhood in the output texture with all the neighborhoods of the same shape in the output texture. This leads to a very high computational load especially if p, the neighborhood size, must be very large to capture the global characteristics of the texture. This is why Paget and Longstaff used a multiscale relaxation, where the information obtained from SR at one resolution is used to constrain the SR at the next highest resolution. By this method, global image characteristics that have been resolved at a low resolution are infused into the relaxation process at the higher resolutions. This helps to reduce the number of iterations required to obtain equilibrium with the Gibbs sampler.

The multigrid representation of an image is shown in Figure 5.3 which is taken from [19]. If $S_0 = [0, M_1] \times [0, M_2]$ represents the pixel's sites of an image x_0, the lower resolutions, or higher grid levels $l > 0$, are decimated versions of the image at level $l = 0$. For a grid level $l > 0$, the image x_l is defined on the lattice $S_l \subset S$, where

$$S_l = \{s = (2^l i, 2^l j)/0 \leq i \leq M_1/2^l, 0 \leq j \leq M_2/2^l\}.$$

The set of sites S_l at level l represents a decimation of the previous set of sites S_{l-1} at the lower grid level $l-1$. The neighborhood system is redefined for each grid level $l > 0$:

$$\mathcal{N}_t^l = \{s \in S_l / \|t - s\|_\infty \leq o\}.$$

For level grid l, SR is not applied to the sites $s \in S^{l+1}$.

We refer to [17] for multiscale representations of Markov random fields. To better incorporate the multiscale relaxation described above, Paget and Longstaff have introduced a pixel temperature function used to determine when to terminate the SR process al one level and start it at the next level. Let l be a grid level. A pixel temperature is incorporated in equation (5.4) by modifying the form of the difference

$$d = y - Y_s. \tag{5.8}$$

In fact at the beginning of the SR at a level l, they define for a site $j \in S_l$ of the output texture the pixel temperature c_j as follows : $c_j = 0$ if $j \in S^{l+1}$ and $c_j = 1$ otherwise. The difference d is replaced

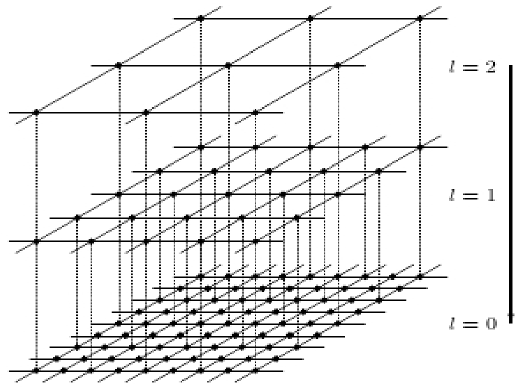


FIG. 5.3: Grid organisation via decimation.

by d' such that :

$$d'_j = (1 - c_{t+j})(x_{t+j} - X_{s+j}), \qquad j \in \mathcal{N}_o.$$

When a pixel x_t has been relaxed in the SR process, we set :

$$\widetilde{T}_t = \max\{0, \frac{\xi + \sum_{j \in \mathcal{N}_o} c_{t+j}}{|\mathcal{N}_o|}\}$$

where $\xi < 0$ is fixed by the user.

Here, the idea is to provide a total confidence to pixels coming from the preceding resolution and to progressively increase the confidence level of a pixel synthesized in the present resolution. When $c_j = 0 \; \forall j \in S_l$, the SR process is considered as having reached an equilibrium state indicating that the image can be propagated to the next lower grid level. This notion of temperature is related to the global temperature used in stochastic annealing (see [12]). Although we have incorporated this pixel temperature function for texture synthesis, we will not study in this paper statistical properties of a such approximation.

5.3.2 Texture synthesis examples

We have incorporated our noncausal bootstrap into the multiscale algorithm with the pixel temperature function described above. Concerning the choice of the parameters :

- For the neighborhood size, we choose $o = 3$ or $o = 4$.
- As in [5], we have not estimate the bandwidth parameter using theoretical results of kernel regression. We have empirically observed that $b = 0.01 \times (\text{neighborhood size})^{1/2}$ provides good results.

– As in [18], we set $\xi = -1$ and generally we have used 4 or 5 grid levels for the synthesis.
Another possibility for the simulation is to use a Conditional Iterative Mode (see [4]). The principle of this deterministic algorithm is to replace each step in the Gibbs sampler by choosing the value X_s such that $F_T\left(dx_t / x_{t+\mathcal{N}_o}\right)$ is maximal or equivalently such that $\|x_{t+\mathcal{N}_o} - Y_s\|$ is minimal in (5.4). Usually this algorithm converges toward a local extremum of the law of the random field on S^{R_T}. This local extremum depends on the initial values put for the pixels on the output texture. We have used the Conditional Iterative Mode for texture synthesis although its definition is not very clear in our case, since the joint laws are not defined.

To illustrate the principle of the multiscale algorithm, Figure 5.4 shows a step of the synthesis in the highest resolution. The Gibbs sampler runs in the raster ordering and Figure 5.4 shows the first sweep. One can see that the lower resolutions give the shape of the texture. Moreover the pixel temperature function helps for a good initialization of the sampler.
This multiscale algorithm does not correctly work only for stochastic textures as in Figure 5.8 but also for highly structured ones as Figure 5.5 shows, even if small discontinuities appear in the last case.
In fact, we have observed that the ICM works as well as the non deterministic algorithm and in some cases better as in Figure 5.6.
Figure 5.9 exhibits a comparison with Efros and Leung's algorithm in a failure case. Texture (b) is taken from Efros and Leung's paper [10] and shows that this causal approach can create garbage when the algorithm slips into a wrong part of the search space. Although the noncausal algorithm does not have the same problem, a gray dark area is often reproduced in texture (c).
Figure 5.10 shows a comparison with two some populars pixel by pixel algorithms. Texture (b) synthesized using Wei et Levoy algorithm [24] and texture (c) using the Ashikhmin method are taken from [2]. The noncausal algorithm used for texture (d) avoids excessive blurring as in (b) and rough images as in (c).

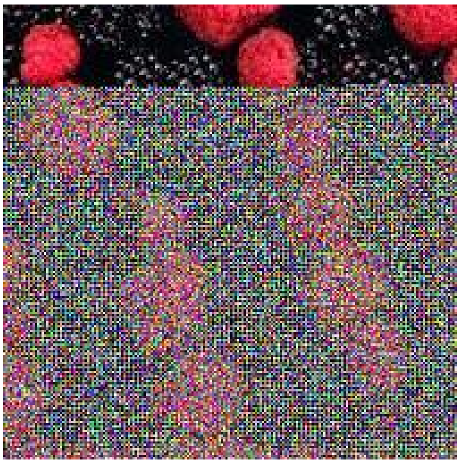
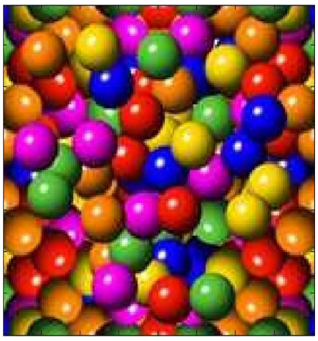
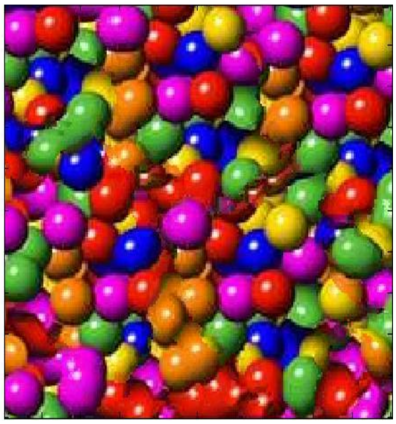


FIG. 5.4: The Gibbs sampler and the highest resolution.

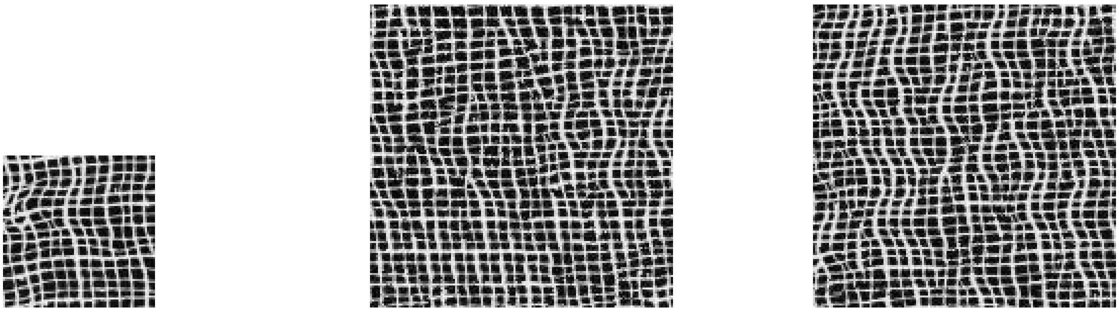


(a)



(b)

FIG. 5.5: (a) Original texture 160×160 pixels, (b) Synthesis with the multiresolution algorithm 200×200 pixels.



(a) (b) (c)

FIG. 5.6: (a) sample 75×75 pixels, (b) multiresolution algorithm 150×150, (c) ICM 150×150.

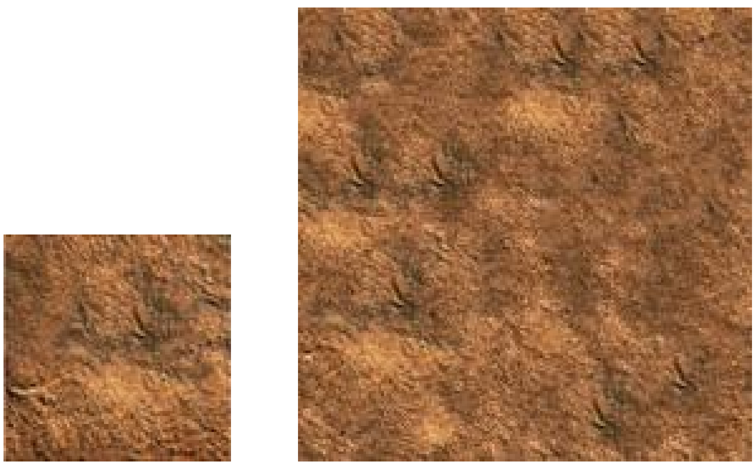


FIG. 5.7: Original texture (128×128) and synthesis (200×200)

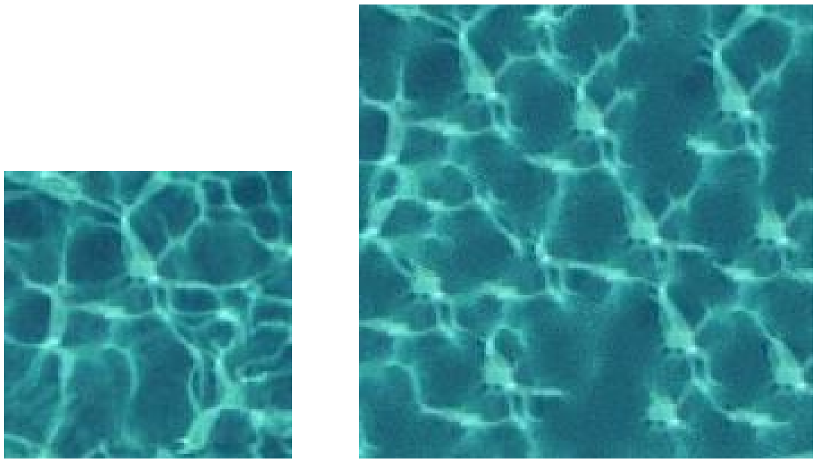


FIG. 5.8: Original texture (128×128) and synthesis (200×200)

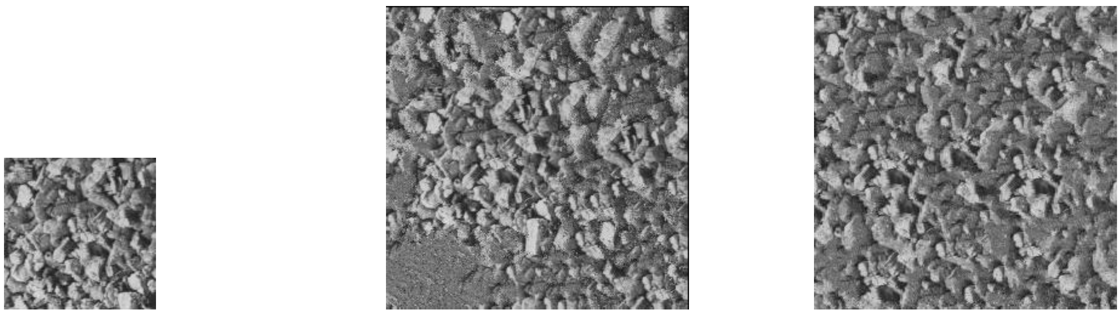


FIG. 5.9: (a) Sample 128×128 pixels, (b) Efros and Leung's result, (c) Our method with $o = 3$ (250×250 pixels in each case).

(a) (b) (c) (d)

FIG. 5.10: Comparison with causal methods : (a) sample 128×128, (b) Wei et Levoy algorithm, (c) Ashikhmin method, (d) ICM. All the synthesized textures are 200×200 pixels.

5.4 Annex

We will use the convenient notation : for $s, t \in \mathbb{N}^* \times \mathbb{N}^*$ define :

$$Y_t = X_{U_t^{(o)}} \text{ and } Y_t(s) = X_{U_t^{(o)}(s)}.$$

Proof of Theorem 5.2

We follow the proof of theorem 2 of Bickel et Levina in order to compute convergence rate. We first recall the following lemma which proof can be found in [8].

Lemma 5.1 *(Moment inequality). Let F_t be a real-valued random field indexed by $I \subset \mathbb{Z}^d$ satisfying conditions* **(A1)**. *If $\mathbb{E}F_t = 0$, $F_t \in \mathbb{L}^{\tau+\varepsilon}$ and $\tau \geq 2$, then there exists a constant C depending only on τ and mixing coefficients of F_t such that*

$$\mathbb{E}\left|\sum_{t\in I} F_t\right|^\tau \leq C \max\left(L(\tau,\varepsilon), L(2,\varepsilon)^{\tau/2}\right),$$

where

$$L(\mu,\varepsilon) = \sum_{t\in I} \left(\mathbb{E}\left|F_t\right|^{\mu+\varepsilon}\right)^{\mu/(\mu+\varepsilon)}.$$

It is easy to see that if $\sup_t \|F_t\|_\infty \leq M$, then we obtain :

$$\mathbb{E}\left|\sum_{t\in I} F_t\right|^\tau \leq CM^\tau \left|I\right|^{\tau/2} \tag{5.9}$$

For $(x,y) \in S \times S^A$, we set :

$$r_T(x,y) = [T]^{-1} \sum_{s\in I_T} \mathbb{1}_{(-\infty,x]}(X_s) W_b(y - Y_s), \quad r(x,y) = \int \mathbb{1}_{(-\infty,x]}(z) f_{X,Y}(z,y)dz,$$

$$f_T(y) = [T]^{-1} \sum_{s\in I_T} W_b(y - Y_s).$$

We have :

$$F_T(x/y) = \frac{r_T(x,y)}{f_T(y)}, \quad F_{X/Y}(x/y) = \frac{r(x,y)}{f_Y(y)}. \tag{5.10}$$

Following the proof of lemma A2 in [5], we prove the following result

Lemma 5.2 *Under assumptions $(A1)-(A4)$, for any $x \in \mathbb{R}$*

$$\sup_{(x,y)\in S\times S^A} \left|r_T(x,y) - r(x,y)\right| = O([T]^{-\gamma})$$

for $0 < \gamma < \frac{\tau-2}{2(v+1)(\tau+v+2)}$.

Proof of Lemma 5.2 In this proof, we will denote by $C > 0$ a generic constant which does not depend on T.
Let $\delta > 0$ such that $b_T = O\left([T]^{-\delta}\right)$, then the proof of lemma A2 in [5] leads to

$$\sup_{(x,y)\in S\times S^A} |\mathbb{E}r_T(x,y) - r(x,y)| = O\left([T]^{-\delta}\right). \tag{5.11}$$

Then we need to bound $\sup_{(x,y)\in S\times S^A} |r_T(x,y) - \mathbb{E}r_T(x,y)|$.
As in [5], we define

$$Z_{t,T}(x,y) = \mathbb{1}_{(-\infty,x]}(X_t)W_{b_T}(y - Y_t) - \mathbb{E}\left(\mathbb{1}_{X_t\le x}W_{b_T}(y - Y_t)\right)$$

and we need to bound $\sup_{(x,y)\in S\times S^A}\left|\frac{1}{[T]}\sum_{t\in I_T} Z_{t,T}(x,y)\right|$.
As $S\times S^A$ is compact, we can cover $S\times S^A$ with N_T cubes $I_{i,T}$ with centers (x_i, y_i) and sides L_T for the supremum norm. Without loss of generality, we suppose $x_1 \le \cdots \le x_{N_T}$ and we set $x_0 = x_1 - L_T$ and $x_{N_T} = x_{N_T} + L_T$. Then

$$\begin{aligned}
\sup_{(x,y)\in S\times S^A}\left|[T]^{-1}\sum_{t\in I_T} Z_{t,T}(x,y)\right| &\le \max_{1\le i\le N_T}\left|[T]^{-1}\sum_{t\in I_T} Z_{t,T}(x_i,y_i)\right| \\
&+ \max_{1\le i\le N_T}\sup_{(x,y)\in(S\times S^A)\cap I_{i,T}}\left|[T]^{-1}\sum_{t\in I_T}(Z_{t,T}(x,y) - Z_{t,T}(x_i,y_i))\right| \\
&= I + II
\end{aligned}$$

– First let deal with term II. Using assumption **(A4)** for the kernel, we have for $t \in I_T$ and $x \in (x_{i-1}, x_i]$:

$$\begin{aligned}
|Z_{t,T}(x,y) - Z_{t,T}(x_i,y_i)| &\le C\left(b_T^{-(v+1)}\|y - y_i\| + b_T^{-v}\left(\mathbb{1}_{x_{i-1}<X_t\le x_i} + \mathbb{P}(x_{i-1} < X_t \le x_i)\right)\right) \\
&\le C\left(b_T^{-(v+1)}L_T + b_T^{-v}\left(\mathbb{1}_{x_{i-1}<X_t\le x_i} + \mathbb{P}(x_{i-1} < X_t \le x_i)\right)\right)
\end{aligned}$$

We choose $L_T = [T]^{-\beta}$ and we set $U_{i,t} = \mathbb{1}_{]x_{i-1},x_i]}(X_t) - \mathbb{P}(x_{i-1} < X_t \le x_{i+1})$. Remark that assumption **(A2)** about the existence of densities allows to derive the bound :

$$\mathbb{P}(x_{i-1} < X_0 \le x_i) \le CL_T.$$

We have :

$$\sup_{(x,y)\in(S\times S^A)\cap I_{i,T}}\left|[T]^{-1}\sum_{t\in I_T}\left(Z_{t,T}(x,y) - Z_{t,T}(x_i,y_i)\right)\right|$$

$$\leq C\left([T]^{\delta(v+1)-\beta}+[T]^{v\delta-1}|\sum_{t\in I_T}U_{i,t}|+[T]^{v\delta}\mathbb{P}(x_{i-1}<X_0\leq x_i)\right)$$

$$\leq C\left([T]^{\delta(v+1)-\beta}+\max_{1\leq i\leq N_T}[T]^{v\delta-1}|\sum_{t\in I_T}U_{i,t}|\right).$$

Now we consider a real number $\gamma<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}$. Since $N_T=O\left(L_T^{-(v+1)}\right)=O\left([T]^{\beta(v+1)}\right)$, we obtain using (A1) and lemma 5.1

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq i\leq N_T}[T]^{v\delta-1}\left|\sum_{t\in I_T}U_{i,t}\right|>[T]^{-\gamma}\right) &\leq \sum_{i=1}^{N_T}[T]^{(\gamma+v\delta-1)\tau}\mathbb{E}\left|\sum_{t\in I_T}U_{i,t}\right|^{\tau}\\ &\leq C[T]^{\beta(v+1)+(\gamma+v\delta-1)\tau+\frac{\tau}{2}}\end{aligned}$$

By the choice of γ, we have $\beta(v+1)+(\gamma+v\delta-1)\tau+\frac{\tau}{2}<-1$ and we deduce from the Borel Cantelli lemma that

$$\max_{1\leq i\leq N_T}[T]^{v\delta-1}\left|\sum_{t\in I_T}U_{i,t}\right|=O\left([T]^{-\gamma}\right)\text{ a.s,}\quad \gamma<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}.$$

Now using the previous inequalities, we deduce that :

$$II\leq O\left([T]^{\delta(v+1)-\beta}+[T]^{-\gamma}\right),\quad \gamma<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}. \tag{5.12}$$

– Now we turn on the term I. For a real number $\widetilde{\gamma}<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}$, we have using (A1) and lemma 5.1 :

$$\begin{aligned}\mathbb{P}(I>[T]^{-\widetilde{\gamma}}) &\leq \sum_{i=1}^{N_T}[T]^{(\widetilde{\gamma}-1)\tau}\mathbb{E}\left|\sum_{t}Z_{t,T}(x_i,y_i)\right|^{\tau}\\ &\leq CN_T[T]^{(\widetilde{\gamma}-1)\tau+\frac{\tau}{2}}b_T^{-v\tau}\\ &\leq C[T]^{\beta(v+1)+(\widetilde{\gamma}-1)\tau+\frac{\tau}{2}+\delta v\tau}\end{aligned}$$

By the choice of $\widetilde{\gamma}$, we have

$$\beta(v+1)+(\widetilde{\gamma}-1)\tau+\frac{\tau}{2}+\delta v\tau<-1,$$

and by the Borel Cantelli lemma, we have

$$I=O\left(T^{-\widetilde{\gamma}}\right)\quad\text{a.s,}\quad \widetilde{\gamma}<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}$$

Now we choose the number β such that :

$$\beta - \delta(v+1) = \frac{\tau - 2 - 2v\delta\tau - 2\beta(v+1)}{2\tau}.$$

This leads to $\beta = \frac{\tau-2+2\tau\delta}{2(v+\tau+1)}$ and to the following rate :

$$I + II = O\left([T]^{-\gamma}\right), \quad \gamma < \frac{\tau - 2 - 2\delta((v+1)^2 + v\tau)}{2(v+\tau+1)}.$$

Finally, we choose δ for an equilibrium with the bound (5.11), solving the equation :

$$\delta = \frac{\tau - 2 - 2\delta((v+1)^2 + v\tau)}{2(v+\tau+1)}.$$

This leads to :

$$\delta = \frac{\tau - 2}{2(v+1)(\tau+v+2)} > 0,$$

which gives the rate given by the Lemma 5.2.
□

Proof of Theorem 5.2 We write :

$$\begin{aligned}
\left|F_T(x/y) - F_{X/Y}(x/y)\right| &= \frac{1}{f_T(y)}\left|r_T(x,y) - r(x,y) + r(x,y) - F_{X/Y}(x/y)f_T(y)\right| \\
&\leq \frac{1}{f_T(y)}\left(\left|r_T(y) - r(y)\right| + F_{X/Y}(x/y)\left|f_T(y) - f_Y(y)\right|\right)
\end{aligned}$$

By lemma 5.2, we have $\sup_{x,y}\left|r_T(x,y) - r(x,y)\right|, \sup_y\left|f_T(y) - f(y)\right| = O\left(T^{-\gamma}\right)$ a.s with $0 < \gamma < \frac{\tau-2}{2(v+1)(\tau+v+2)}$. Since by (A2), $\inf_{y\in S^A} f_Y(y) > 0$ and $\sup_{(x,y)\in S\times S^A} F_{X/Y}(x/y) \leq 1$, we get the result.□

5.5 Some tools for the consistency : Continuity results

In this section, we give two results which describe the behavior of random fields in relation to their one point conditional distributions. If it is possible to construct a nonparametric estimate of the one point conditional distribution which is also compatible with the one point conditional distribution of a Markov random field, the following results will be useful to describe the statistical properties of the joint laws of the model.
We first give a lemma which states the behavior of the joint laws of a sequence of random fields when their one-point conditional distributions are convergent. In the sequel, let S be a compact set of $\mathbb{R}$

endowed with its Borelian algebra $\mathcal{B}(S)$. Let $\mathcal{X} = S^I$ where I is a denumerable set. For any sequence $(u_i)_{i\in I}$ of positive real numbers satisfying $\sum_{i\in I} u_i < \infty$, we consider the distance d on $\mathcal{X}$ defined by :

$$d(z,z') = \sum_{i\in I} u_i \left|z_i - z_i'\right|, \qquad z,z' \in \mathcal{X}.$$

Then $(\mathcal{X}, d)$ is a compact metric space. For $A \subset I$, let $p_A : \mathcal{X} \to S, \quad z \mapsto z_A$. For $t \in I$, we will write p_t instead of $p_{\{t\}}$. Moreover, for $t \in I$, we set :

$$\mathcal{F}_t^- = \sigma(p_j / \quad j \neq t).$$

We denote by $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$. If ν_1, ν_2 are two elements of $\mathcal{P}(\mathcal{X})$, the Prohorov distance d_P between ν_1 and ν_2 is defined by :

$$d_P(\nu_1,\nu_2) = \inf\{\varepsilon > 0, \nu_1(A) \leq \nu_2(A^\varepsilon) + \varepsilon, \forall A \in \mathcal{B}(\mathcal{X})\},$$

where $A^\varepsilon = \{z \in \mathcal{X} / d(z,A) \leq \varepsilon\}$. The distance d_P defines the weak convergence on $\mathcal{P}(\mathcal{X})$ which is a compact space topology.
Now for $\nu \in \mathcal{P}(\mathcal{X})$ and any bounded measurable function f on $\mathcal{X}$, we set :

$$\mathbb{E}_\nu(f) = \int f d\nu.$$

For $t \in I$, we denote ν_t the kernel on $\mathcal{P}(\mathcal{X})$ such that :

$$\nu_t(A/z) = \mathbb{E}_\nu\left(\mathbb{1}_A / \mathcal{F}_t^-\right)(z),$$

where $\mathbb{E}_\nu(\cdot/\mathcal{F})$ denotes the conditional expectation with respect to a $\sigma-$algebra $\mathcal{F} \subset \mathcal{B}(\mathcal{X})$.

Finally let $\gamma = (\gamma_t)_{t\in I}$ be a sequence of probability kernels such that for $t \in I$, γ_t is a kernel from $\mathcal{F}_t^-$ to $\mathcal{B}(\mathcal{X})$ satisfying the property :

$$\gamma_t\left(A \cap B/\cdot\right) = \gamma_t\left(A/\cdot\right) \times \mathbb{1}_B, \quad (A,B) \in \sigma(p_t) \times \mathcal{F}_t^-.$$

If h is a bounded measurable function on $\mathcal{X}$, we denote $\gamma_t(h)$ the measurable function on $\mathcal{X}$ such that :

$$\gamma_t(h)(z) = \int f(w)\gamma_t(dw/z), \qquad z \in \mathcal{X}.$$

We define the following subset of $\mathcal{P}(\mathcal{X})$:

$$\mathcal{G}(\gamma) = \left\{\nu \in \mathcal{P}(\mathcal{X}) / \quad \forall t \in I, \quad \nu\gamma_t = \nu\right\},$$

where for all $(t,\nu) \in I \times \mathcal{P}(\mathcal{X})$, $\nu\gamma_t$ denotes the element of $\mathcal{P}(\mathcal{X})$ such that :

$$\nu\gamma_t(A) = \int \gamma_t(A/z) d\nu(z).$$

We will say that γ satisfies the condition **(C)** if :

$$(\mathbf{C}) \qquad \forall t \in I, \quad h \in \mathcal{C}(\mathcal{X}) \Rightarrow \gamma_t(h) \in \mathcal{C}(\mathcal{X}),$$

where $\mathcal{C}(\mathcal{X})$ is the space of continuous and bounded functions on $\mathcal{X}$.
The following result gives the behavior of a sequence of random fields in the case of uniform convergence of their one point conditional distribution. A general treatment of topological properties of random fields is given in [13]. For completeness of this work, we state and prove the following result :

Theorem 5.3 *For $t \in I$, let γ_t be a probability kernel on $\mathcal{X} \times \mathcal{F}_t^-$. Suppose that the sequence $\gamma = (\gamma_t)_{t\in I}$ satisfies condition* **(C)**. *Then for a sequence $(\nu^{(n)})_n$ of $\mathcal{P}(\mathcal{X})$ such that*

$$\sup_{(x,z)\in S\times\mathcal{X}} \left|\nu_t^{(n)}(p_t \leq x/z) - \gamma_t(p_t \leq x/z)\right| \to_{n\to\infty} 0,$$

we have $d_P(\nu^{(n)}, \mathcal{G}(\gamma)) \to_{n\to\infty} 0$.

Proof of Lemma 5.3 Suppose that there exists $\varepsilon > 0$ and a subsequence $s = (\nu^{(n_k)})_{k\in\mathbb{N}}$ such that $d_P(\nu_{n_k}, \mathcal{G}(\gamma)) > \varepsilon$, $\forall k \in \mathbb{N}$. Since this sequence is relatively compact with respect to the weak topology, then there exists a subsequence $(\nu^{(n_{k'})})_{k'}$ of s and $\nu \in \mathcal{P}(\mathcal{X})$ such that :

$$\lim_{k'\to\infty} \nu^{(n_{k'})} = \nu.$$

We are going to show that $\nu \in \mathcal{G}(\gamma)$, which is a contradiction with $d_P(\nu^{n_{k'}}, \mathcal{G}(\gamma)) > \varepsilon$, $k' \in \mathbb{N}$.
Let $h \in C(\mathcal{X})$. For $t \in I$, we have :

$$\begin{aligned}
& \int h d\nu^{(n_{k'})} - \int h d(\nu\gamma_t) \\
= & \int \left(\nu_t^{(n_{k'})}(h) - \gamma_t(h)\right) d\nu^{(n_{k'})} + \int \gamma_t(h) d\nu^{(n_{k'})} - \int \gamma_t(h) d\nu \\
= & A + B.
\end{aligned}$$

Let $\varepsilon > 0$. Since h is uniformly continuous on $\mathcal{X}$, there exists $\delta > 0$ such that $d(z, z') < \delta \Rightarrow |h(z) - h(z')| < \varepsilon$. We choose a subdivision $a_1, \ldots, a_k$ of the interval $[a, b] \supset S$ with step smaller then δ/u_t. Let h_k the function defined on $[a, b]$ by $h_k(z) = \sum_{l=1}^{k-1} h(z(l))\mathbb{1}_{]a_l,a_{l+1}]}(z_0)$ where for $l = 1, \ldots, k-1$, $z(l)$ is the element of $\mathcal{X}$ such that $z(l)_t = a_l$ and $z(l)_s = z_s$ if $s \neq t$. We have $\sup_{z\in\mathcal{X}} |h(z) - h_k(z)| < \varepsilon$.
We deduce :

$$\left|\nu_t^{(n_{k'})}(h)(z) - \gamma_t(h)(z)\right| \leq 2\varepsilon + 2(k-1)\|h\|_\infty \sup_{(x,z)\in S\times\mathcal{X}} \left|\nu_t^{(n)}(p_t \leq x/z) - \gamma_t(p_t \leq x/z)\right|.$$

One can conclude that :

$$\sup_{z \in \mathcal{X}} \left| \nu_t^{(n_{k'})}(h)(z) - \gamma_t(h)(z) \right| \rightarrow_{k' \rightarrow \infty} 0.$$

Hence $A \rightarrow 0$.

Moreover since $\gamma_t(h) \in C(\mathcal{X})$ by condition **(C)**, the weak convergence of the sequence $\left(\nu^{(n_{k'})}\right)_{k' \in \mathbb{N}}$ implies $B \rightarrow 0$.

Then we conclude that for $t \in I$, we have $\nu\gamma_t = \nu$, and the result follows from this contradiction.$\square$

Now we investigate the following problem. Suppose for simplicity that $S = [a, b]$. Assume that for $\mu, \nu \in \mathcal{P}(\mathcal{X})$, the distance between the conditional distribution functions $\nu_t(p_t \leq \cdot / \cdot)$ and $\mu_t(p_t \leq \cdot / \cdot)$ is known, then is it possible to obtain the distance between μ and ν over some cylinders sets of the form

$$C_{x_{t_1}, \dots, x_{t_k}} = \otimes_{i=1}^{k} [a, x_{t_i}] \times S^{I \setminus \{t_1, \dots, t_k\}}?$$

In other words, can we obtain the distance between the distribution functions of the joint laws ? This problem is linked to the phase transition phenomenon and to the Dobrushin's contraction formula. In order to apply this formula, the following assumption will be needed :

(H) We assume that there exist two families of non negative real numbers $\{L_{t,j}/t, j \in I\}$ and $\{M_{t,j}/t, j \in I\}$, with $L = \sup_{t \in I} \sum_{j \neq t} L_{t,j} < 1$ and $M = \sup_{t \in I} \sum_{j \neq t} M_{t,j} < \infty$, such that $\forall z, z' \in \mathcal{X}$:

$$\sup_{x \in S} \left| \mu\left(p_t \leq x/z\right) - \mu\left(p_t \leq x/z'\right) \right| \leq \sum_{j \neq t} M_{t,j} \left| z_j - z'_j \right|,$$

$$\int_S \left| \mu\left(p_t \leq x/z\right) - \mu\left(p_t \leq x/z'\right) \right| dx \leq \sum_{j \neq t} L_{t,j} \left| z_j - z'_j \right|.$$

Theorem 5.4 *Let $\mu, \nu \in \mathcal{P}(\mathcal{X})$ with $S = [a, b]$. Suppose that the random field μ satisfies assumption* **(H)**. *Then for each finite subset $\{t_1, \dots, t_k\}$ of I, we have :*

$$\sup_{(x_{t_1}, \dots, x_{t_k}) \in S^k} \left| \mu\left(C_{x_{t_1}, \dots, x_{t_k}}\right) - \nu\left(C_{x_{t_1}, \dots, x_{t_k}}\right) \right| \leq C \sup_{t \in I} \sup_{(x,z) \in S \times \mathcal{X}} \left| \mu(p_t \leq x/z) - \nu(p_t \leq x/z) \right|,$$

where $C = 1 + M(b-a)\left(k - 1 + \frac{1}{1-L}\right)$.

We first recall the following inequality due to Dobrushin (see [11] remark 2.17). This inequality allows to bound the distance between two random fields μ and ν with the distance between their local conditional specification. Of course, a such inequality implies that there is no phase transition. Some contraction conditions on the local conditional specifications are needed to get this inequality. Before giving this inequality in our case, we introduce some notations. Let r be a metric on $S = [a, b]$. If α and β are two probability on S, the Warsserstein metric is defined as

$$R(\alpha, \beta) = \sup \frac{\left| \int f d\alpha - \int f d\beta \right|}{\delta(f)},$$

where the supremum is taken over all Lipshitz functions f on S with

$$\delta(f) = \sup_{x \neq x'} \frac{|f(x) - f(x')|}{r(x, x')} < \infty.$$

For our result, we will only consider the metric $r(x, x') = |x - x'|$. One can mention that in this case, R has the following expression :

$$R(\alpha, \beta) = \int |\alpha([a, x]) - \beta([a, x])| \, dx, \quad \alpha, \beta \in \mathcal{P}(S). \tag{5.13}$$

We define also $L(\mathcal{X})$ the space of real functions f such that :

$$\left| f(z) - f(z') \right| \leq \sum_{i \in \mathbb{Z}^2} \left| z_i - z'_i \right| \delta_i(f), \quad \sum_{i \in \mathbb{Z}^2} \delta_i(f) < \infty$$

where

$$\delta_i(f) = \sup_{z \neq z'} \left\{ \frac{|f(z) - f(z')|}{|z_i - z'_i|} / z_j = z'_j, \forall j \neq i \right\}.$$

Let $\mu, \nu \in \mathcal{P}(\mathcal{X})$. We suppose that the following continuity condition holds :

$$f \in L(\mathcal{X}) \Rightarrow \forall t \in \mathbb{Z}^2, \mu_t(f) \in L(\mathcal{X}). \tag{5.14}$$

For $\mu \in \mathcal{P}(\mathcal{X})$, the contraction coefficients are defined by

$$C_{ik} = \sup \left\{ \frac{R\left(\mu(p_k \in \cdot / z), \mu(p_k \in \cdot / z')\right)}{|z_i - z'_i|} / z_j = z'_j, \forall j \neq i \right\}$$

and

$$b_k = \int R\left(\mu(p_k \in \cdot / z), \nu(p_k \in \cdot / z)\right) \nu(dz)$$

Note that with the expression (5.13), we have the bound :

$$b_k \leq (b - a) \sup_{(x,z) \in S \times \mathcal{X}} \left| \mu(p_k \leq x / z) - \nu(p_k \leq x / z) \right|. \tag{5.15}$$

Let $D = \sum_{n \geq 0} C^n$ where C^n denotes the nth power of the matrix C. D is well defined for example if

$$c = \sup_{k \in \mathbb{Z}^2} \sum_{i \in \mathbb{Z}^2} C_{ik} < 1. \tag{5.16}$$

In this case, the following inequality holds :

$$\left| \int f d\mu - \int f d\nu \right| \leq \sum_{i \in \mathbb{Z}^2} (bD)_i \delta_i(f), \quad f \in L(\mathcal{X}). \tag{5.17}$$

Proof of Theorem 5.4 First, from the point 1 of lemma 5.3, condition (5.14) is satisfied for μ. Moreover for $i,k \in \mathbb{Z}^2$, we have $C_{i,k} = L_{k,i}$ and the condition (6.9) is satisfied with $c = L$. Then, inequality (5.17) holds for ν.

For $l \in \{1,\ldots,k\}$ and $x \in \mathcal{X}$, we set :

$$f_l(z) = \prod_{m=1}^{l} \mathbb{1}_{(-\infty, x_{t_m}]}(z_{t_m})$$

and

$$g_l = \mu_{t_l} \circ \cdots \circ \mu_{t_1}(f_l), \quad h_l = \nu_{t_l} \circ \cdots \circ \nu_{t_1}(f_l).$$

We have :

$$\begin{aligned} \left|\int f_k d\mu - \int f_k d\nu\right| &= \left|\int g_k d\mu - \int h_k d\nu\right| \\ &\leq \left|\int g_k d\mu - \int g_k d\nu\right| + \left|\int g_k d\nu - \int h_k d\nu\right| \\ &\leq \left|\int g_k d\mu - \int g_k d\nu\right| + \beta_k \end{aligned}$$

where $\beta_l = \|g_l - h_l\|_\infty$, $l \in \{1,\ldots,k\}$.
First by lemma 5.4, one can apply inequality (5.17) to the function g_k. We obtain :

$$\left|\int g_k d\mu - \int g_k d\nu\right| \leq \sum_{i\in I} (bD)_i \delta_i(g_k).$$

Using for $i \in I$, the inequality $\sum_{j\in I} C_{j,i}^n \leq c^n = L^n$ and the bound given in (5.15), we have :

$$(bD)_i \leq \sum_{j\in I} D_{j,i} \sup_{j\in I} b_j \leq \frac{b-a}{1-L} \sup_{t\in I} \sup_{(x,z)\in S\times\mathcal{X}} \left|\mu(p_t \leq x/z) - \nu(p_t \leq x/z)\right|, \tag{5.18}$$

Then using lemma 5.4, we conclude that :

$$\left|\int g_k d\mu - \int g_k d\nu\right| \leq M\frac{b-a}{1-L} \sup_{t\in I} \sup_{(x,z)\in S\times\mathcal{X}} \left|\mu(p_t \leq x/z) - \nu(p_t \leq x/z)\right|.$$

If we use the bound for β_k in Lemma 5.5, we conclude that :

$$\left|\int f_k d\mu - \int f_k d\nu\right| \leq C \sup_{t\in I} \sup_{(x,z)\in S\times\mathcal{X}} \left|\mu(p_t \leq x/z) - \nu(p_t \leq x/z)\right|,$$

with $C = 1 + M(b-a)\left(k - 1 + \frac{1}{1-L}\right)$. The proof of theorem 5.4 is now complete. □

Lemma 5.3 *Let $g \in L(\mathcal{X})$ and $t \in I$. Then*

1. *$\mu_t(g) \in L(\mathcal{X})$ and :*
$$\sum_{i\in I} \delta_i(\mu_t(g)) \leq \sum_{i\in I} \delta_i(g).$$
2. *$\|\mu_t(g) - \nu_t(g)\|_\infty \leq (b-a)\delta_t(g) \sup_{(x,z)\in S\times\mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)|.$*

Proof of lemma 5.3

1. For $(x,z,t) \in S \times \mathcal{X} \times \mathbb{Z}^2$, with $(xz)_t$ the element r of $\mathcal{X}$ such that $r_t = x$ and $r_s = z_s$ if $s \neq t$. Let $g_{z,t} : S \to \mathbb{R}$, $x \to g((xz)_t)$ is a Lipshitz function and then is derivable almost everywhere with $g'_{z,t}$ satisfying $\|g'_{z,t}\|_\infty \leq \delta_t(g)$. Let $z, \widetilde{z} \in \mathcal{X}$. With an integration by parts formula, we have :
$$\begin{aligned} A &= \left| \int g_{\widetilde{z},t}(x) \left(\mu(p_t \in dx/z) - \mu(p_t \in dx/\widetilde{z})\right) \right| \\ &= \left| \int \left(\mu(p_t < dx/z) - \mu(p_t < x/\widetilde{z})\right) g'_{z,t}(x) dx \right| \\ &\leq \delta_t(g) \sum_{i\neq t} L_{t,i} \left| z_i - \widetilde{z}_i \right|. \end{aligned}$$
Now this leads to :
$$\begin{aligned} & \left| \int g_{z,t}(x)\mu(p_t \in dx/z) - \int g_{\widetilde{z},t}(x)\mu(p_t \in dx/\widetilde{z}) \right| \\ \leq & \sum_{i\neq t} \delta_i(g) \left| z_i - \widetilde{z}_i \right| + A \\ \leq & \sum_{i\neq t} \delta_i(g) \left| z_i - \widetilde{z}_i \right| + \delta_t(g) \sum_{i\neq t} L_{t,i} \left| z_i - \widetilde{z}_i \right| \end{aligned}$$
From this bound, it is obvious that $\mu_t(g) \in L(\mathcal{X})$ if $g \in L(\mathcal{X})$ since :
$$\sum_{i\in I} \delta_i(\mu_t(g)) \leq \sum_{i\neq t} \delta_i(g) + L\delta_t(g) < \infty.$$
2. For $z \in \mathcal{X}$, with another integration by parts formula :
$$\begin{aligned} |\mu_t(g)(z) - \nu_t(g)(z)| &= \left| \int g_{z,t}(x)\mu(p_t \in dx/z) - \int g_{z,t}(x)\nu(p_t \in dx/z) \right| \\ &\leq \left| \int \left(\mu(p_t < x/z) - \nu(p_t < x/z)\right) g'_{z,t}(x) dx \right| \\ &\leq (b-a)\delta_t(g) \sup_{(x,z)\in S\times\mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)| \square \end{aligned}$$

Lemma 5.4 *For $l \in \mathbb{N}^*$, we have $g_l \in L(\mathcal{X})$ and :*
$$\sum_{i\in I} \delta_i(g_l) \leq M.$$

Proof of Lemma 5.4 First using assumption **(A5)**, $g_1 \in L(\mathcal{X})$ and we have :

$$\sum_{i \in i} \delta_i(g_1) \leq \sum_{i \neq t_1} M_{t_1,i} \leq M.$$

Using the point 1 of lemma 5.3, a straightforward finite induction shows that $g_l \in L(\mathcal{X})$, $l \leq k$. Now, if $l \geq 2$, with the point 1 in lemma 5.3 and the definition of g_l,

$$\sum_{i \in I} \delta_i(g_l) \leq \sum_{i \in I} \delta_i(g_{l-1}),$$

and Lemma 5.4 follows.□

Lemma 5.5 *For $l \in \mathbb{N}^*$, we have :*

$$\beta_k \leq (1 + (k-1)(b-a)M) \times \max_{t \in \{t_1,\dots,t_k\}} \sup_{(x,z) \in S \times \mathcal{X}} \left| \mu(p_t \leq x/z) - \nu(p_t \leq x/z) \right|.$$

Proof of Lemma 5.5 We have by definition $\beta_1 \leq \sup_{(x,z) \in S \times \mathcal{X}} \left| \mu(p_{t_1} \leq x/z) - \nu(p_{t_1} \leq x/z) \right|.$ Now, let $l \in \{1, \dots, k-1\}$. We have using the point 2 of Lemma 5.3 and Lemma 5.4 :

$$\begin{aligned}
\beta_{l+1} &= \left\| \mu_{t_{l+1}}(g_l) - \nu_{t_{l+1}}(h_l) \right\|_\infty \\
&\leq \left\| \nu_{t_{l+1}}(g_l) - \nu_{t_{l+1}}(h_l) \right\|_\infty + \left\| \nu_{t_{l+1}}(g_l) - \mu_{t_{l+1}}(g_l) \right\|_\infty \\
&\leq \beta_l + (b-a)\delta_{t_{l+1}}(g_l) \times \sup_{(x,z) \in S \times \mathcal{X}} \left| \mu(p_{t_{l+1}} \leq x/z) - \nu(p_{t_{l+1}} \leq x/z) \right| \\
&\leq \beta_{T,l} + (b-a)M \sup_{(x,z) \in S \times \mathcal{X}} \left| \mu(p_{t_{l+1}} \leq x/z) - \nu(p_{t_{l+1}} \leq x/z) \right|
\end{aligned}$$

One can easily deduce the result.□

Bibliographie

[1] Abend, K., Harley, T. J. and Kanal, L. N. (1965) *Classification of binary random patterns* IEEE Trans. Inform. Theory ITŨ11 538Ũ544.

[2] Ashikhmin, M. (2001) *Synthesizing Natural Textures.* The proceedings of 2001 ACM Symposium on Interactive 3D Graphics, Research Triangle Park, NorthCarolina March 19-21, pp. 217-226.

[3] De Bonet, J.S (1997) *Multiresolution sampling procedure for analysis and synthesis of texture images.* In Proc. 24th Annual Conference on Computer Graphics and Interactive Techniques 361-368. ACM Press, New York.

[4] Besag, J. (1986) *On the Statistical Analysis of Dirty Pictures.* J. R. Statis. Soc., Series B, Vol. 48, No. 3, pp. 259-302.

[5] Bickel, P., Levina, E. (2006) *Texture synthesis and nonparametric resampling of random fields* The Annals of Statistics, Vol. 34, No 4, 1751-1773.

[6] Cross, G. C., Jain, A. K. (1983) *Markov random field texture models* IEEE Transactions on Pattern Analysis and Machine Intelligence, 5(1) :25-39.

[7] Dehardt, J. (1971) *Generalizations of the Glivenko-Cantelli theorem.* Ann. Math. Statist., 2050-2055.

[8] Doukhan, P. (1994) *Mixing : Properties and Examples.* Lecture Notes in Statist. 85. Springer, New York. MR1312160

[9] Efros, A. A., Freeman, W. T. (2001) *Image quilting for texture synthesis and transfer.* In Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques 341-346. ACM Press, New York.

[10] Efros, A. A., Leung, T. (1999) *Texture synthesis by non-parametric sampling.* In Proc. IEEE International Conference on Computer Vision 2 1033-1038. IEEE Computer Soc., Washington.

[11] Föllmer, H. (1988) *Random Fields and Diffusion Processes.* Ecole dŠEté de Probabilités de St. Flour XVI. Lecture Notes in Mathematics 1362, Springer, 101-203.

[12] Geman, S., Geman, D. (1984) *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images.* IEEE Trans. Pattern Anal. Mach. Intell, 6, 721-741, 1984.

[13] Georgii, H.-O. (1988) *Gibbs measures and phase transitions.* de Gruyter, Berlin.

[14] Guyon, X. (1995) *Random fields on a network.* Springer, Berlin.

[15] Heeger, D.J., Bergen, J. R. (1995) *Pyramid-based texture analysis/synthesis.* In Proc. 22nd Annual Conference on Computer Graphics and Interactive Techniques 229-238. ACM Press, New York.

[16] Liang, L., Liu, C., Xu, Y., Guo, B., Shum, H.-Y. (2001) *Real-time texture synthesis by patch-based sampling.* Technical Report MSR-TR-2001-40, Microsoft Research.

[17] Luettgen, M.R., Karl, W.C, Wilsky, A.S, Tenney, R.R. (1993) *Multiscale Representations of Markov Random fields.* IEEE transactions on signal processing, vol. 41, nř 12, 3377-3396.

[18] Paget, R. (1999) *Nonparametric Markov random field models for natural texture images.* PhD Thesis, University of Queensland.

[19] Paget, R., Longstaff, I. D. (1998) *Texture synthesis via a noncausal nonparametric multiscale Markov random field* IEEE Transactions on Image Processing, vol. 7, n 6, pp. 925-931.

[20] Paparoditis, E., Politis, D. (2002) *The local bootstrap for Markov processes.* J. Statist. Plann. Inference, nř 108, pp. 301-328.

[21] Popat, K., Picard, R. (1993) *Novel cluster-based probability model for texture synthesis, classification, and compression.* In Proc. SPIE Visual Communications and Image Processing 2094 756-768. SPIE, Bellingham, WA.

[22] Rajarshi, M. (1990) *Bootstrap in Markov sequences based on estimates of transition density.* Ann. Inst. Math. Statist. 42 253-268.

[23] Simoncelli, E., Portilla, J. (1998) *Texture characterization via joint statistics of wavelet coefficient magnitudes.* In Proc. Fifth IEEE International Conference on Image Processing 1, 62-66. IEEE Computer Soc., Washington.

[24] Wei, L.-Y., Levoy, M. (2000) *Fast texture synthesis using tree-structured vector quantization.* In Proc. 27th Annual Conference on Computer Graphics and Interactive Techniques 479-488. ACM Press, New York.

[25] Zhu, S. C., Wu, Y. N., Mumford, D. (1998) *Filters, random fields and maximum entropy (FRAME)- toward a unified theory for texture modeling.* Internat. J. Computer Vision 27, 107-126.

9

Chapitre 6

An integer-valued bilinear type model

Abstract

A integer-valued bilinear type model is proposed. It can take positive as well as negative values. The existence of the process is established in $\mathbb{L}^m$. In fact, this process is the unique causal solution to an equation that is similar to a classical bilinear type model equation. For the estimation of the parameters, we suggest a quasi-maximum likelihood approach. The estimator is strongly consistent and asymptotically normal.

Note

The content of this part is based on a paper, written in collaboration with Alain Latour.

6.1 Introduction

As pointed in [19], integer-valued times series are common in practice. In epidemiology, we often consider the number of cases of a given disease over a 28-day period. In this context, the data are collected to make sure that the population is not threatened by an epidemic. As well as in intensive care monitoring, where vital parameters have to be analyzed online, good modeling is required. As soon as three consecutive values seem to be too high, governmental actions are planed to avoid the widespread of the disease, since there may be serious consequences for the population otherwise. See [12] where regression methods are used to perform intensive care monitoring.
Concerning integer-valued time series, we may refer the reader to [9, 16, 17, 19]. For a review of various models and their statistical properties, we do recommend [9] where some extensions of integer autoregressive and moving average models are also presented. Many models encountered in the literature are based on thinning operators as defined in [22]. In this paper, we use a more general definition.

Definition 6.1 *Let $Y = (Y_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed (independent and identically distributed) integer-valued random variables with mean α independent of an integer-valued variable X. The thinning operator, $\alpha\circ$ is defined by :*

$$\alpha \circ X = \begin{cases} sign(X) \sum_{i=1}^{|X|} Y_i, & \text{if } X \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The sequence $(Y_i)_{i\in\mathbb{N}}$ is referred to as a counting sequence. This definition is more general than the usual one where Y is a sequence of Bernoulli random variables with expected value α. (See, for example, [6].) Here, it is a sequence of i.i.d non-negative integer-valued variables, for example, a sequence of Poisson distributed variables Y_i with parameter α. In fact, any non-negative integer-valued random sequence can be used as a counting series. More, X can take negative values.
To avoid any confusion, if necessary, we can denote the operator by $\alpha(Y)\circ$ or $\alpha(\theta)\circ$ instead of $\alpha\circ$ to clearly indicate that it is based on the sequence Y or that it depends on the parametric vector θ of the distribution of the variables involved in the operator. Nevertheless, we prefer the simplest notation.
The reader should bear in mind that in Definition 6.1, the mean of the summands Y_i associated with the operator $\alpha\circ$ is α. Suppose $\tilde{\alpha}\circ$ is another thinning operator based on a counting sequence $\left(\widetilde{Y}_i\right)_{i\in\mathbb{N}}$.
The operators $\alpha\circ$ and $\tilde{\alpha}\circ$ are said to be independent if, and only if, the counting sequences $(Y_i)_{i\in\mathbb{N}}$ and $\left(\widetilde{Y}_i\right)_{i\in\mathbb{N}}$ are mutually independent.

Example 6.1 (Branching process with immigration) *The Bienaymé-Galton-Watson (BGW) process with immigration can be written using a thinning operator. With this notation, if the offspring of an individual is distributed as Y, and if ζ_t is the immigration contribution to the population at the t^{th} generation, then the classical BGW process satisfies*

$$X_t = \alpha \circ X_{t-1} + \zeta_t. \tag{6.1}$$

For each generation t, we need a counting sequence Y_t, so $(Y_t)_{t\in\mathbb{Z}}$ is an i.i.d process of i.i.d sequences $(Y_{t,j})_{j\in\mathbb{N}}$. In the case of a BGW process, X_t is never negative. The links between branching processes with immigration and $INARMA(p, q)$ is clearly identified and explained in [2]. □

Example 6.2 (Inventory monitoring) *Suppose X_t represents the number of widgets remaining in a distributor inventory at the end of a month. Also suppose if the distributor runs out of stock, he registers the customer order to send it as soon as the widget becomes available. In that case the number of items left at the end of the month could be negative.* □

Example 6.3 *Given two counting processes, $(X_t)_t$ and $(Y_t)_t$, in some situations we may be interested in the difference between the two processes : $Z_t = X_t - Y_t$, $t \in \mathbb{Z}$, is the excess of X_t over Y_t. Clearly, Z_t can be negative.*

It is clear that in many situations, standard univariate models are not appropriate in the context of integer-valued time series analysis. Using classical real-valued models is even more critical when we cope with a low frequency count data. This has been pinpointed by [21] and many more authors (see [11, 20]). It could explain why integer-valued processes are an important topic and why there have been so many papers on the subject for more than twenty-five years. Many authors use thinning operators to define integer-valued process similar to classical econometric models. See, for examples, [2, 3, 6, 8, 13].

In [21], a worthy discussion is made on integer-valued ARMA(p, q) processes. In the latter paper, an efficient MCMC algorithm is presented for a wide class of integer-valued autoregressive moving-average processes. In many papers, p and q are assumed to be known. In [7], efficient order selection algorithms are studied for these integer-valued ARMA processes.

It is clear that integer-valued ARMA processes cannot satisfy all practitioner expectations. A common working hypothesis is that the observed time series comes from a stationary process. In some situations, there are good reasons to doubt about this hypothesis.

For example, in Figure 6.1, we give X_t, the number of campylobacteriosis[1] cases in the Northern Québec, starting in January 1990, with an observation every 28 days.

One may believe that $\mathbb{E}X_t$ increases with t. Also, perhaps that there is a structural change happening in the neighborhood of the 100^{th} observation. In [10], problems with this series are clearly identified. For reasons that are similar to the ones we met when we tackle the problem of modeling real valued time series, we have to develop well-adapted tools for practitioner needs. A Dickey-Fuller unit-root type test has been studied by [14]. For a GARCH type model, [8] suggested a process with Poisson conditional distribution with mean and variance λ_t. In [3], the authors tackled the problem of an integer-valued bilinear process. They restricted their works to the following model :

$$X_t = a \circ X_{t-1} + b \circ (\varepsilon_{t-1} X_{t-1}) + \varepsilon_t,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence of non-negative integer-valued random variables. They proved the existence of this stochastic process, suggested appropriate estimators under a Poissonian hypothesis and applied it to a social medicine series. Recently, [5] cleverly proved the existence of a more general version of this process and inspired in this paper our existence proof of another process (see (6.8)).

The paper has the following structure. In Section 6.2, we recall a result from [4] giving conditions for the existence of a solution to a quite general model equation in which X_t is expressed in terms of its own past values and the present and past values of a sequence of independent and identically distributed random variables (*cf.* (6.4)). A quite simple approximation $\left(X_t^{(n)}\right)_t$ to $(X_t)_t$ is also given. For this approximation we have :

$$X_t^{(n)} \to X_t \quad \text{in } \mathbb{L}^m \quad \text{and} \quad X_t^{(n)} \to X_t \quad \text{a.s.}.$$

[1] Infection with a Campylobacter species is one of the most common causes of human bacterial gastroenteritis.

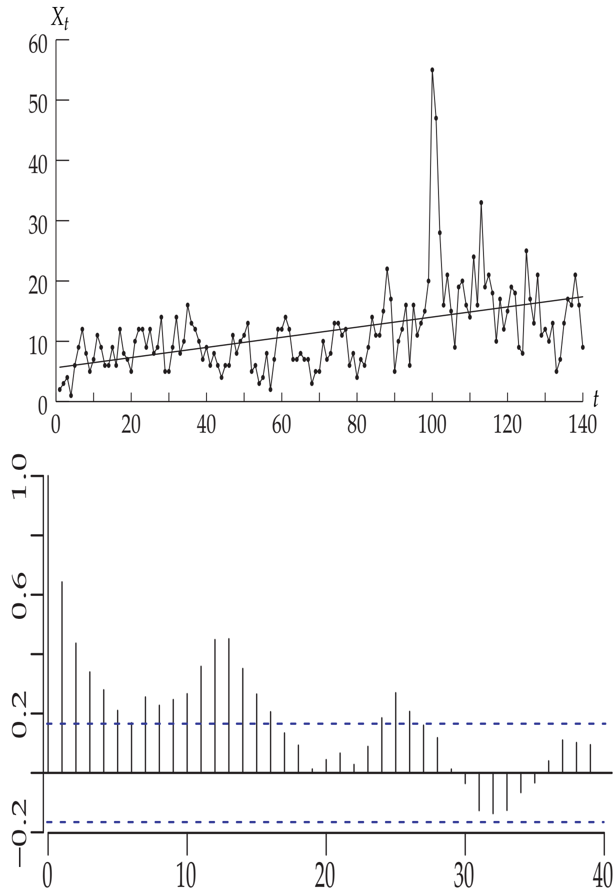


FIG. 6.1: Number of campylobacteriosis cases in the Northern Québec, starting in January 1990, 13 regular observations per year. To the top, is the graphic of the original series, to the bottom, is the sample simple correlogram.

In Section 6.3, we give some basic properties of Definition 6.1 thinning operators. Then, two models are presented : the INLARCH model and an integer-valued bilinear type model. Simple conditions for the existence of these processes are given.

Section 6.4 is devoted to estimation of the parameters. The problem is tackled using a quasi-maximum likelihood estimator for the bilinear model parameters. Before announcing the properties of the estimator, working assumptions and hypotheses are enunciated. Theorem 6.4 claims the strong consistency of the estimators and Theorem 6.1 gives its asymptotic distribution.

In Section 6.5, we comment consequences of the results when we consider the *almost* classical GINAR(p) process. Proofs are postponed to Section 6.6.

6.2 The model

From now on, the sequence $(\xi_t)_{t\in\mathbb{Z}}$ is i.i.d and takes values in a space E' (in many cases E' is just $\mathbb{R}^\infty$). Let $(E, \|\cdot\|)$ be a Banach space. For a random variable $Z \in E$ and a real number $m \geqslant 1$, the expression $\|Z\|_m$ stands for $(\mathbb{E}\|Z\|^m)^{1/m}$ and $E^{(\mathbb{N})}$, a subset of $E^{\mathbb{N}}$, denotes the set of sequences in $E^{\mathbb{N}}$ with a finite number of non-null terms. Let $F : E^{(\mathbb{N})} \times E' \to E$ be a measurable function and assume there exists a sequence of functions $(a_j)_{j\in\mathbb{N}}$ such that for all $(x_j)_{j\in\mathbb{N}}$ and $(y_j)_{j\in\mathbb{N}}$ in $E^{(\mathbb{N})}$,

$$\|F(0,0,\ldots;\xi_0)\|_m < +\infty, \tag{6.2a}$$

$$\|F(x_1,x_2,\ldots;\xi_0) - F(y_1,y_2,\ldots;\xi_0)\|_m \leqslant \sum_{j=1}^{\infty} a_j \|x_j - y_j\|, \tag{6.2b}$$

with

$$\sum_{j=1}^{\infty} a_j := a < 1. \tag{6.3}$$

Let us recall a general result of [4] about existence and approximation in $\mathbb{L}^m$ of a stationary process $(X_t)_{t\in\mathbb{Z}}$, solution of (6.4) :

$$X_t = F(X_{t-1}, X_{t-2}, \ldots; \xi_t). \tag{6.4}$$

The following theorem is a consequence of Theorem 3 and Lemma 6 of [4].

Theorem 6.1 *Assume properties (6.2a) and (6.2b) hold for some $m \geqslant 1$, then there exists a unique stationary solution of (6.4) such that*

$$X_t \in \sigma(\xi_t, \xi_{t-1}, \ldots), \qquad t \in \mathbb{Z}. \tag{6.5}$$

Moreover, the sequence of stationary processes defined $\forall t \in \mathbb{Z}$ as

$$X_t^{(n)} = \begin{cases} F(0;\xi_0), & n = 0; \\ F(X_{t-1}^{(n-1)}, X_{t-2}^{(n-1)}, \ldots; \xi_t), & n \geqslant 1; \end{cases}$$

satisfies

$$X_t^{(n)} \to X_t \quad \textit{in } \mathbb{L}^m \quad \textit{and} \quad X_t^{(n)} \to X_t \quad \textit{a.s.}.$$

Remark 6.1 *A solution of (6.4) which satisfies (6.5) is always ergodic. Indeed from (6.5), we have :*

$$\bigcap_{t\in\mathbb{Z}} \sigma(X_{t-1}, X_{t-2}\ldots) \subset \bigcap_{t\in\mathbb{Z}} \sigma(\xi_{t-1}, \xi_{t-2}\ldots) \tag{6.6}$$

As ξ is i.i.d, any event in the σ-field of the right-hand side of (6.6) has probability 0 or 1 from which we conclude that any event in the σ-field of the left-hand side is also of probability 0 or 1. This shows that the process $(X_t)_{t\in\mathbb{Z}}$ is ergodic. The argument comes from [6].

In the sequel, a solution (6.4) satisfying (6.5) will be called a causal solution in $\mathbb{L}^m$. Note that a such solution implies the independence of the σ-algebras $\sigma(X_u : u \leq s)$ and $\sigma(\xi_v : v \geq t)$ when $t > s$.

6.3 Construction of integer-valued models

6.3.1 Basic properties of signed thinning operators

Lemma 6.1 *Lets X, Z two random variables and Y, $\widetilde{Y}$ two counting sequences associated with the operators $\alpha\circ$ and $\tilde{\alpha}\circ$, respectively. Suppose the variance of the counting sequence variables are β and $\tilde{\beta}$, respectively. Assume that (X, Z), Y, $\widetilde{Y}$ are independent. Let $m \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$. Then :*

1. $\mathbb{E}\alpha \circ X = \alpha \mathbb{E}X$ *and* $\mathbb{E}(\alpha \circ X)^2 = \beta \mathbb{E}|X| + \alpha^2 \mathbb{E}X^2$.
2. $\mathbb{E}(\alpha \circ X)(\tilde{\alpha} \circ Z) = \alpha\tilde{\alpha}\mathbb{E}XZ$ *and* $Cov(\alpha \circ X, \tilde{\alpha} \circ Z) = \alpha\tilde{\alpha}Cov(X, Z)$.
3. $\|\alpha \circ X - \alpha \circ Z\|_m \leq \|Y\|_m \|X - Z\|_m$.
4. *For $l \geq 2$, we have*
$$\|\alpha \circ X\|_l \leq |\alpha| \, \|X\|_l + c_l \, \|Y - \alpha\|_l \, \|X\|_{l-1}^{1/2}$$
where the constant $c_l > 0$ only depends on l.

Remark 6.2 *Consider the simple model :*

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t. \tag{6.7}$$

For $t \in \mathbb{Z}$, let $\xi_t = \big((Y_{t,i})_{i\in\mathbb{N}^*}, \varepsilon_t\big)$. For $x \in \mathbb{Z}$, we define :

$$F(x, \xi_0) = \alpha \circ x_1 + \varepsilon_0.$$

Suppose $(\xi_t)_t$ is an i.i.d sequence and let $m = 2$. From the result 3. of Lemma 6.1, one has :

$$\|F(x; \xi_0) - F(y; \xi_0)\|_2 \leq \|Y\|_2 \, |x - y| \, .$$

Moreover if $F(0; \xi_0) = \varepsilon_0 \in \mathbb{L}^2$, we can apply Theorem 6.1 if $\|Y\|_2 < 1$. But this is not optimal. Indeed, it is well known that the condition $\alpha < 1$ is sufficient for the existence and uniqueness in $\mathbb{L}^m$ of a stationary solution of (6.7) (see [18]).
This is the reason why the construction of the model is in two steps. Firstly, we apply Theorem 6.1 with $m = 1$ and get a solution in $\mathbb{L}^1$. Then we use a contraction condition on the means of the counting sequences. Secondly, we show that this solution is still unique in $\mathbb{L}^m$, m being an integer.

6.3.2 Bilinear model

Let $(X_t)_{t\in\mathbb{Z}}$ be a solution to the equation :

$$X_t = \sum_{j=1}^{\infty} \alpha_j \circ X_{t-j} + \varepsilon_t \left(\sum_{j=1}^{\infty} \beta_j \circ X_{t-j} \right) + \eta_t, \tag{6.8}$$

where η_t and ε_t are integer valued random variables in $E = \mathbb{Z}$, $\alpha_j\circ$ and $\beta_j\circ$ being signed thinning operators associated with counting sequences $Y^{(j)}$ and $\widetilde{Y}^{(j)}$ respectively. Theorem 6.2 gives conditions for the existence of a solution to (6.8).
Suppose $\mathbb{E}\varepsilon_t = 0$ and for each $t \in \mathbb{Z}$, let :

$$\xi_t = \left(\left(Y_{t,i}^{(j)} \right)_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*}, \left(\widetilde{Y}_{t,i}^{(j)} \right)_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*}, \varepsilon_t, \eta_t \right).$$

The random variable ξ_t takes values in $\mathbb{Z}^{\mathbb{N}^*\times\mathbb{N}^*} \times \mathbb{Z}^{\mathbb{N}^*\times\mathbb{N}^*} \times \mathbb{Z} \times \mathbb{Z}$. We suppose the process $(\xi_t)_{t\in\mathbb{Z}}$ is i.i.d

Theorem 6.2 *Suppose for an integer $m \geq 1$,*

$$a = \sum_{j=1}^{\infty} \left\| Y^{(j)} \right\|_1 + \|\varepsilon_0\|_m \left\| \widetilde{Y}^{(j)} \right\|_1 < 1, \qquad \sum_{j=1}^{\infty} \left\| Y^{(j)} \right\|_m + \left\| \widetilde{Y}^{(j)} \right\|_m + \|\eta_0\|_m < \infty, \tag{6.9}$$

then there exists a unique causal solution to (6.8) in $\mathbb{L}^m$.

6.3.3 INLARCH(∞) time series model

An INLARCH(∞) time series model satisfies

$$X_t = \alpha \circ \varepsilon_t + \sum_{j=1}^{\infty} \alpha_j \circ (\varepsilon_t X_{t-j}), \quad t \in \mathbb{Z}. \tag{6.10}$$

For $j \in \mathbb{N}^*$, we will denote by $Y^{(j)}$ (resp. Y) the counting sequences associated with the operator $\alpha_j\circ$ (resp. $\alpha\circ$).
As for the bilinear model, we suppose $(\xi_t)_{t\in\mathbb{Z}}$ is an i.i.d sequence. Theorem 6.3 states a sufficient condition for the existence of this process.

Theorem 6.3 *Suppose for an integer $m \geq 1$,*

$$a = \|\varepsilon\|_m \sum_{j\in\mathbb{N}^*} \left\| Y^{(j)} \right\|_1 < 1, \qquad \sum_{j\in\mathbb{N}^*} \left\| Y^{(j)} \right\|_m + \|Y\|_m < \infty,$$

then equation (6.10) admits a unique causal solution in $\mathbb{L}^m$.

6.4 Quasi-maximum likelihood estimator in bilinear model

This section aims at giving a quasi-maximum likelihood estimators (QMLE) for the parameters of the bilinear model (6.8) with a finite number of terms in the two summations. Without lost of generality, we may assume that there are p terms in each summation ; otherwise some $\alpha_j\circ$ or $\beta_j\circ$ are $0\circ$. The equation satisfied by the process is :

$$X_t = \sum_{j=1}^{p} \alpha_j \circ X_{t-j} + \varepsilon_t \sum_{j=1}^{p} \beta_j \circ X_{t-j} + \eta_t. \tag{6.11}$$

Example Let $p = 2$ and consider

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-1} + \varepsilon_t \beta_1 \circ X_{t-1} + \eta_t,$$

where : $\alpha_1\circ$ is based on a Bernoulli counting series with $p = 1/2$; $\alpha_2\circ$ and $\beta_1\circ$, on Poisson counting series with means 1/8 and 1/2, respectively ; $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d Poisson random variables with parameter $\lambda = 1/2$; $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of differences between two independent Poisson variables with parameter $\lambda = 2/3$. A simulated trajectory is presented in Figure 6.2. Note that there is "period", just after $t = 80$ with quite high values compared to the other ones.

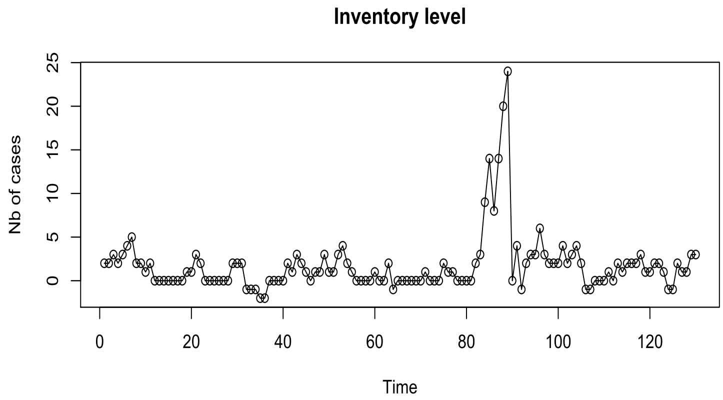


FIG. 6.2: Simulated trajectory generated by model $X_t = \alpha_1\circ X_{t-1}+\alpha_2\circ X_{t-1}+\varepsilon_t\beta_1\circ X_{t-1}+\eta_t$, where : $\alpha_1\circ$ is based on a Bernoulli counting series with $p = 1/2$; $\alpha_2\circ$ and $\beta_1\circ$, on Poisson counting series with means 1/8 and 1/2, respectively ; $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d Poisson random variables with parameter $\lambda = 1/2$; $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of differences between two independent Poisson variables with parameter $\lambda = 2/3$.

For $(t, j) \in \mathbb{Z} \times \{1, \ldots, p\}$, we define the following $\sigma-$algebras :

$$\mathcal{F}_t = \sigma(X_{t-k} : k \in \mathbb{N}^*), \quad \mathcal{G}_{t,j} = \sigma(Y_{t,i}^{(j)} : i \in \mathbb{N}^*), \quad \text{and} \quad \widetilde{\mathcal{G}}_{t,j} = \sigma(\widetilde{Y}_{t,i}^{(j)} : i \in \mathbb{N}^*),$$

From now on, we suppose the following working assumptions are satisfied :

1. $(\xi_t)_{t\in\mathbb{Z}}$ is an i.i.d sequence of random variables.
2. For all $t \in \mathbb{Z}$, the σ-algebras $\mathcal{G}_{t,1}, \ldots, \mathcal{G}_{t,p}$ (resp. $\widetilde{\mathcal{G}}_{t,1}, \ldots, \widetilde{\mathcal{G}}_{t,p}$) are independent.
3. For all $t \in \mathbb{Z}$, the σ-algebras $\sigma(\varepsilon_t)$, $\sigma(\eta_t)$ and $(\vee_{1\leq j\leq p}\mathcal{G}_{t,j}) \vee (\vee_{1\leq j\leq p}\widetilde{\mathcal{G}}_{t,j})$ are mutually independent.

Remark 6.1 *Assumptions with respect to the σ-algebras allow dependence between the set of operators $\{\alpha_j\circ\}_{1\leqslant j\leqslant p}$ and the set of operators $\{\beta_j\circ\}_{1\leqslant j\leqslant p}$.*

For an integer $d \geq 1$, let θ be a subset of $\mathbb{R}^d$ and $\theta_0 \in \theta$. For $1 \leq j \leq p$, consider functions $b_j, c_j, w_j, \mu, \nu : \theta \to \mathbb{R}$ such that :

i) $b_j(\theta_0) = \alpha_j$ and $c_j(\theta_0) = \beta_j$.
To ensure identifiability, we suppose there exists $j_0 \in \{1, \ldots, p\}$ such that $\beta_{j_0} > 0$ and the function c_{j_0} is positive on θ.

ii) $w_j(\theta_0) = \operatorname{Var} Y^{(j)} + \sigma^2 \times \operatorname{Var} \widetilde{Y}^{(j)}$, $\sigma^2 = \operatorname{Var} \varepsilon_0$.

iii) $\mu(\theta_0) = \mathbb{E}\eta_0$ and $\nu(\theta_0) = \operatorname{Var} \eta_0$.

The following hypotheses will also be required.

H1) Θ is a compact subset of $\mathbb{R}^d$.

H2) Condition (6.9) holds with $m = 2$.

H3) The distribution support of η_t contains at least 5 different points if $\operatorname{Var} \varepsilon_t \neq 0$ and 3, otherwise.

H4) The following condition is satisfied : $h = \inf_{\theta \in \theta} \nu(\theta) > 0$.

H5) The function $f : \theta \to \mathbb{R}^{3p+2}$ defined by

$$f(\theta) = \Big(\left(b_j(\theta), c_j(\theta), w_j(\theta)\right)_{1 \leq j \leq p}, \mu(\theta), \nu(\theta) \Big)$$

is injective and continuous on θ.

For $(t, \theta) \in \mathbb{Z} \times \theta$, let

$$m_t(\theta) = \mu(\theta) + \sum_{j=1}^{p} b_j(\theta) X_{t-j}$$

and

$$V_t(\theta) = \sigma^2 \left(\sum_{j=1}^{p} c_j(\theta) X_{t-j} \right)^2 + \sum_{j=1}^{p} w_j(\theta) \left| X_{t-j} \right| + \nu(\theta).$$

Observe that under assumption H4, we have :

$$\inf_{\theta \in \theta} V_t(\theta) \geq h, \qquad \text{a.s.} \tag{6.12}$$

Lemma 6.2 *Let $(X_t)_{t \in \mathbb{Z}}$ given by (6.11). We have :*

$$\mathbb{E}\left(X_t / \mathcal{F}_{t-1}\right) = m_t(\theta_0), \quad \mathit{Var}\left(X_t / \mathcal{F}_{t-1}\right) = V_t(\theta_0).$$

Remark 6.2 *On the one hand, the conditional expectation is the same as the one of a GINAR(p) process. On the other hand, a second-degree polynomial appears in the conditional variance.*

6.4.1 Estimators definition

For the estimation of the parameters, no distribution assumptions are made and a quasi-maximum likelihood approach turns out to be well suited to this setup. The maximum is found assuming a conditional Gaussian density for X_t, given the past until time $t-1$. In [23] this method is used in ARCH modeling.
Let us give the details for model (6.11). Suppose we observe $X_0, \dots, X_{-p+1}$ and let :

$$\begin{aligned}
q_t(\theta) &= \frac{(X_t - m_t(\theta))^2}{V_t(\theta)} + \ln V_t(\theta), \qquad t \geq 1; \\
Q_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} q_t(\theta); \\
Q(\theta) &= \mathbb{E}\left(\frac{(X_0 - m_0(\theta))^2}{V_0(\theta)} + \ln V_0(\theta)\right); \\
\hat{\theta}_T &= \arg\min_{\theta\in\Theta} Q_T(\theta).
\end{aligned}$$

So, $\hat{\theta}_T$ is the QMLE for θ and the actual value of θ is θ_0.

Consistency of the estimator.

Theorem 6.4 *Under hypotheses H1 to H5, the estimator $\hat{\theta}_T$ is a strongly consistent estimator of θ_0 : $\lim_{T\to\infty}\hat{\theta}_T = \theta_0$ a.s.*

Asymptotic normality of QMLE.

In the following, if g is a function, $g : \theta \mapsto \mathbb{R}$, ∇g is its gradient and $\nabla^2 g$ is its Hessian matrix.

Other hypotheses are needed.

H7) Condition (6.9) holds with $m = 4$.

H8) The f function is twice differentiable on Θ and $rank\nabla f(\theta_0) = d$. More, $\inf_{\theta\in\Theta} w_j(\theta) > 0, \forall j = 1, \dots, p$.

H9) θ_0, the actual value of θ, is an interior point of Θ, θ_0, that is $\theta_0 \in \Theta^{\circ}$.

Theorem 6.1 *Under hypotheses H1, ..., H9, the estimator $\hat{\theta}_T$ is asymptotically normal :*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \to_{T\to\infty} \mathcal{N}(0, F_0^{-1}G_0F_0^{-1}) \quad \textit{in distribution},$$

where

$$\begin{aligned} F_0 &= \mathbb{E}\nabla^2 q_0(\theta_0) \\ &= \mathbb{E}\left(V_0(\theta_0)^{-2}\nabla V_0(\theta_0)\nabla V_0(\theta_0)'\right) + 2\mathbb{E}\left(V_0(\theta_0)^{-1}\nabla m_0(\theta_0)\nabla m_0(\theta_0)'\right) \end{aligned}$$

and

$$\begin{aligned} G_0 &= Var\,\nabla q_0(\theta_0) \\ &= \mathbb{E}\left(V_0(\theta_0)^{-4}(X_0 - m_0(\theta_0))^4\nabla V_0(\theta_0)\nabla V_0(\theta_0)'\right) \\ &\quad - \mathbb{E}\left(V_0(\theta_0)^{-2}\nabla V_0(\theta_0)\nabla V_0(\theta_0)'\right) + 4\mathbb{E}\left(V_0(\theta_0)^{-1}\nabla m_0(\theta_0)\nabla m_0(\theta_0)'\right) \\ &\quad + \mathbb{E}\left(V_0(\theta_0)^{-3}(X_0 - m_0(\theta_0))^3\nabla V_0(\theta_0)\nabla m_0(\theta_0)'\right) \\ &\quad + \mathbb{E}\left(V_0(\theta_0)^{-3}(X_0 - m_0(\theta_0))^3\nabla m_0(\theta_0)\nabla V_0(\theta_0)'\right) \end{aligned}$$

6.5 QMLE for GINAR(p) processes

When $\sigma^2 = 0$, (6.11) leads to a GINAR(p) process :

$$X_t = \sum_{j=1}^{p} \alpha_j \circ X_{t-j} + \eta_t.$$

Estimation for this process has been tackled by least squares (see [6, 18]).
The conditional least squares estimator is given by :

$$\hat{\theta}_T = \arg\min_{\theta\in\theta} \frac{1}{T}\sum_{t=1}^{T}\left(X_t - m_t(\theta)\right)^2.$$

But this approach cannot be applied to obtain estimators for all the parameters if the probability distribution of the counting sequences depends on two parameters or more.
In particular, suppose the operators $\alpha_j\circ$, $1 \leqslant j \leq p$, are counting series with variables for which the support is a 3-point set $\{a, b, c\}$ and we want to estimate :

$$(q_{a,j}, q_{b,j}) = \left(\mathbb{P}\left(Y_{0,0}^{(j)} = a\right), \mathbb{P}\left(Y_{0,0}^{(j)} = b\right)\right)$$

$1 \leq j \leq p$, as well as $(\mathbb{E}\eta_0, \mathrm{Var}\,\eta_0)$. Let

$$\theta_0 = (q_{a,1}, q_{b,1}, \ldots, q_{a,p}, q_{b,p}, \mathbb{E}\eta_0, \mathrm{Var}\,\eta_0) \in \Theta \subset \mathbb{R}^{2p+2}.$$

For $1 \leq j \leq p$, let

$$b_j(\theta) = (a - c)\theta_{2j-1} + (b - c)\theta_{2j} + c$$

$$w_j(\theta) = (a^2 - c^2)\theta_{2j-1} + (b^2 - c^2)\theta_{2j} + c^2 - b_j(\theta)^2$$

$$\mu(\theta) = \theta_{2p+1}, \qquad \nu(\theta) = \theta_{2p+2}$$

A least squares approach is not tractable because θ_0 is not identifiable by just considering $m_t(\theta) = \sum_{j=1}^{p} b_j(\theta)X_{t-j}$, $t \in \mathbb{Z}$, the conditional means of the process $(X_t)_{t\in\mathbb{Z}}$. In fact, the function $\theta \mapsto (b_1(\theta), \ldots, b_p(\theta), \mu(\theta), \nu(\theta))$ is not injective. However, it is clear that the function :

$$\theta \mapsto (b_1(\theta), w_1(\theta), \ldots, b_p(\theta), w_p(\theta), \mu(\theta), \nu(\theta))$$

is injective and we can use Section 6.4.1 results to estimate parameter θ_0.

Example 6.4 *Let us return to Example 6.4. It is quite easy to proceed to the estimation of the parameters using a widespread and simple tool like Microsoft Excel. We use the Excel's Solver macro to find the optimum. The estimated model is :*

$$X_t = \hat{\alpha}_1 \circ X_{t-1} + \hat{\alpha}_2 \circ X_{t-1} + \varepsilon_t \hat{\beta}_1 \circ X_{t-1} + \eta_t,$$

where : $\hat{\alpha}_1\circ$ is based on a Bernoulli counting series with $\hat{p} = 0.65$; $\hat{\alpha}_2\circ$ and $\hat{\beta}_1\circ$, on Poisson counting series with means 0.12 and 0.58 respectively ; $(\eta_t)_t$ a sequence of i.i.d Poisson random variables with parameter $\lambda = 0.47$; $(\varepsilon_t)_t$ is a sequence of differences between two independent Poisson variables with parameter $\lambda = 0.49$. So, $\hat{\theta} = (0.645; 0.120, 0.503; 0.669; 0.469)'$. Recall that the actual value of the parameter is : $\theta = (0.5; 0.125; 0.5; 0.66\overline{7}; 0.500)'$.

6.6 Extended proofs of the results

6.6.1 Proof of Lemma 6.1

1. $\mathbb{E}\left(\alpha \circ X/X = x\right) = \mathbb{E}\left(sign(x)\sum_{i=1}^{|x|} Y_i\right) = x\alpha$ and the first result follows from expectation with respect to X.

   For the second point, note that :

   $$\mathbb{E}\left((\alpha \circ X)^2/X = x\right) = \mathbb{E}\left(\sum_{i=1}^{|x|} Y_i^2\right) = |x|\,\mathbb{E}Y^2 + |x|\,(|x| - 1)\alpha^2,$$

   and again the result follows from expectation with respect to X.

2. Since the variables $(\alpha \circ X, Z)$ and $\widetilde{Y}$ are independent, from result 1., we get the following equality :

   $$\mathbb{E}\left((\alpha \circ X) \times (\widetilde{\alpha} \circ Z)\right) = \widetilde{\alpha}\mathbb{E}\left((\alpha \circ X) \cdot Z\right).$$

   As Y is independent of (X, Z), we obtain $\widetilde{\alpha}\mathbb{E}\left((\alpha \circ X) \cdot Z\right) = \alpha\widetilde{\alpha}\mathbb{E}XZ$. The second assertion is obvious.

3. We use the first point of item 1 and if $x,\ z \in \mathbb{Z}$:

$$\|\alpha \circ x - \alpha \circ z\|_m \leq \|Y\|_m |x - z| .$$

Independence between Y and (X, Z) yields the result after expectation with repesct to X and Z.

4. See [5, Theorem 2.2], for a proof of this inequality.

6.6.2 Proof of Theorem 6.2

The demonstration proceeds in two steps. Firstly, we show that under Theorem 6.2 hypotheses, equation (6.8) has a unique causal solution in $\mathbb{L}^1$. Then, we show that this solution has moments of order m.
To show the existence in $\mathbb{L}^1$, we use Theorem 6.1. Let $F : \mathbb{Z}^{(\mathbb{N}^*)} \times \mathbb{Z} \to \mathbb{Z}$ be :

$$F\left((x_j)_{j\in\mathbb{N}^*};\xi_0\right) = \sum_{j=1}^{\infty} \alpha_j \circ x_i + \varepsilon_0 \left(\sum_{j=1}^{\infty} \beta_j \circ x_i\right) + \eta_0.$$

We have : $\|F(0;\xi_0)\|_1 = \|\eta_0\|_1 < \infty$. More, by the result 3. of Lemma 6.1, we get :

$$\begin{aligned}\left\|F\left((x_j)_{j\in\mathbb{N}^*};\xi_0\right) - F\left((y_j)_{j\in\mathbb{N}^*};\xi_0\right)\right\|_1 &\leq \sum_{j=1}^{\infty} \left(\left\|Y^{(j)}\right\|_1 + \|\varepsilon_0\|_1 \left\|\widetilde{Y}^{(j)}\right\|_1\right) |x_j - y_j| \\ &\leq \sum_{j=1}^{\infty} \left(\left\|Y^{(j)}\right\|_1 + \|\varepsilon_0\|_m \left\|\widetilde{Y}^{(j)}\right\|_1\right) |x_j - y_j| .\end{aligned}$$

Because $a = \sum_{j=1}^{\infty} \left(\left\|Y^{(j)}\right\|_1 + \|\varepsilon_0\|_m \left\|\widetilde{Y}^{(j)}\right\|_1\right) < 1$, we can apply Theorem 6.1 and conclude that there exists in $\mathbb{L}^1$ a unique causal stationary process $(X_t)_{t\in\mathbb{Z}}$, solution to (6.8), such that $\|X_t\|_1 < \infty$.
Let us show that $X_t \in \mathbb{L}^m$. To this end, let us introduce the stationary process defined by :

$$X_{n,t} = \begin{cases} F(0;\xi_t), & n = 0; \\ F(\{X_{n-1,t-j}\}_{j\geq 1};\xi_t), & n \geqslant 1; \end{cases} \quad t \in \mathbb{Z}.$$

By Theorem 6.2, we have :

$$X_{n,t} \to_{n\to\infty} X_t \quad \text{a.s and} \quad X_{n,t} \to_{n\to\infty} X_t \quad \text{in } \mathbb{L}^1.$$

Next, we show that $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_m < \infty$. From this last inequality, using Fatou's Lemma, we will conclude that

$$\|X_0\|_m \leq \liminf_{n\to\infty} \|X_{n,0}\|_m < \infty.$$

We use induction to show that for each $l \in \{1,\dots,m\}$, we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_l < \infty$. Since $\lim_{n\to\infty} X_{n,0} = X_n$ in $\mathbb{L}^1$, we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_1 < \infty$ and the result follows for $l=1$. Suppose for $l \in \{1,\dots,m-1\}$ we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_l < \infty$. We want to show that $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_{l+1} < \infty$. Let $n \in \mathbb{N}$. We have :

$$\|X_{n+1,0}\|_{l+1} \le \sum_{j=1}^{\infty} \|\alpha_j \circ X_{n,-j}\|_{l+1} + \|\varepsilon_0\|_{l+1} \sum_{j=1}^{\infty} \|\beta_j \circ X_{n,-j}\|_{l+1} + \|\eta_0\|_{l+1}.$$

To simplify the equations writing, let :

$$d_{j,h} = \left\|Y^{(j)} - \alpha_j\right\|_h + \|\varepsilon_0\|_h \left\|\widetilde{Y}^{(j)} - \beta_j\right\|_h, \quad \text{for } j \ge 1 \text{ and } h \in \{1,\dots,m\}.$$

Using result 4. of Lemma 6.1, we get :

$$\begin{aligned}
\|X_{n+1,0}\|_{l+1} &\le c_{l+1} \sum_{j=1}^{\infty} d_{j,l+1} \|X_{n,-j}\|_l^{1/2} + \|\eta_0\|_{l+1} \\
&+ \sum_{j=1}^{\infty} \left(|\alpha_j| + |\beta_j| \|\varepsilon_0\|_{l+1}\right) \|X_{n,-j}\|_{l+1} \\
&\le a \|X_{n,0}\|_{l+1} + B.
\end{aligned}$$

where $B = c_{l+1} \sup_k \|X_{k,0}\|_l^{1/2} \sum_{j=1}^{\infty} d_{j,l+1} + \|\eta_0\|_{l+1}$.
As $X_{0,0} = \eta_0$, this leads to $\|X_{n+1,0}\|_{l+1} \le a^{n+1} \|\eta_0\|_{l+1} + B \sum_{i=1}^{n} a^i$.
Observe that

$$B \le c_{l+1} \sup_k \|X_{k,0}\|_l^{1/2} \sum_{j=1}^{\infty} d_{j,m} + \|\eta_0\|_m.$$

Then by condition (6.9) and the induction hypothesis, B is finite and we get :

$$\|X_{n+1,0}\|_{l+1} \le \|\eta_0\|_{l+1} + \frac{B}{1-a}$$

and $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_{l+1} < \infty$. Hence, by finite induction on the subset $\{1,\dots,m\}$, we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_m < \infty$. Finally, by the remark made previously, $\|X_0\|_m < \infty$. Uniqueness in $\mathbb{L}^m$ follows from uniqueness in $\mathbb{L}^1$.

6.6.3 Proof of Lemma 6.2

The conditional expectation of X_t given the past until time $t-1$ is :

$$\mathbb{E}\left(X_t/\mathcal{F}_{t-1}\right) = \mathbb{E}\eta_0 + \sum_{j=1}^{p} \mathbb{E}\left(\alpha_j \circ X_{t-j}/\mathcal{F}_{t-1}\right) = \mathbb{E}\left(\eta_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}\right) = m_t(\theta_0).$$

For the conditional variance, we get :

$$\mathrm{Var}\ (X_t/\mathcal{F}_{t-1}) = \mathrm{Var}\left(\sum_{j=1}^{p} \alpha_j \circ X_{t-j}/\mathcal{F}_{t-1}\right) + \sigma^2 \mathbb{E}\left(\sum_{j=1}^{p} \beta_j \circ X_{t-j}/\mathcal{F}_{t-1}\right) + \mathrm{Var}\,\eta_0$$

Simple computations lead to

$$\mathbb{E}\left((\alpha_j \circ X_{t-j})(\alpha_k \circ X_{t-k})/\mathcal{F}_{t-1}\right) = \begin{cases} \alpha_j \alpha_k X_{t-j} X_{t-k}, & j \neq k; \\ \alpha_j^2 X_{t-j}^2 + \mathrm{Var}\,Y^{(j)} \left|X_{t-j}\right|, & j = k. \end{cases}$$

Similar formulas can be found if $\beta_j \circ$ is substituted for $\alpha_j \circ$. Using these expressions in Var $(X_t/\mathcal{F}_{t-1})$ expansion leads to the final expression :

$$\mathrm{Var}\ (X_t/\mathcal{F}_{t-1}) = \sigma^2 \left(\sum_{j=1}^{p} \beta_j X_{t-j}\right)^2 + \sum_{j=1}^{p} w_j(\theta_0) \left|X_{t-j}\right| + \mathrm{Var}\,\eta,$$

where $w_j(\theta_0) = \mathrm{Var}\,Y^{(j)} + \sigma^2 \mathrm{Var}\,\widetilde{Y}^{(j)}$. This is exactly $V_t(\theta_0)$.

6.6.4 Proof of Theorem 6.3

The proof is very similar to proof given to Theorem 6.2 and is omitted.

6.6.5 Proof of Theorem 6.4

Before giving the demonstration, some intermediate results are required. Let us recall Theorem 6.5 from [23].

Theorem 6.5 *Let Θ a compact set of $\mathbb{R}^d$ and $(v_t)_{t\in\mathbb{Z}}$ a stationary ergodic sequence of random elements with values in $\mathcal{C}(\Theta, \mathbb{R})$. Then the uniform strong law of large numbers is implied by*

$$\mathbb{E} \sup_{\theta\in\Theta} |v_0(\theta)| < \infty.$$

Lemma 6.3 follows from Theorem 6.5.

Lemma 6.3

$$\sup_{\theta\in\Theta} |Q_T(\theta) - Q(\theta)\| \rightarrow_{T\rightarrow\infty} 0.$$

Proof. Let us verify that Theorem 6.5 hypotheses are satisfied. Firstly, we prove that $\forall\theta \in \theta$, $\{q_t(\theta)\}_{t\geqslant 1}$ is an ergodic stationary sequence. From Remark 6.1, $(X_t)_{t\in\mathbb{Z}}$ is a stationary ergodic process. More, for $(t, \theta) \in \mathbb{Z} \times \Theta$, by definition of $q_t(\theta)$, there exists a measurable function f_θ defined on $\mathbb{R}^{p+1}$ such that $q_t(\theta) = f_\theta\left(X_t, \ldots, X_{t-p}\right)$. This implies that the sequence $\{q_t(\theta)\}_{t\in\mathbb{Z}}$ is also stationary and ergodic.

Secondly, we have :

$$|q_0(\theta)| \leq \frac{[X_0 - m_0(\theta)]^2}{h} + |\ln(V_0(\theta))| .$$

Then, from H1, H2, H4 and H5, we get :

$$|X_0 - m_0(\theta)| \leq |X_0| + \|\mu\|_\infty + \sum_{j=1}^{p} \|b_j\|_\infty |X_{-j}| \in \mathbb{L}^2,$$

and

$$h \leq V_0(\theta) \leq \sigma^2 \left(\sum_{j=1}^{p} \|c_j\|_\infty |X_{-j}| \right)^2 + \sum_{j=1}^{p} \|w_j\|_\infty |X_{-j}| + \|\nu\|_\infty \in \mathbb{L}^1.$$

This shows that $\mathbb{E} \sup_{\theta \in \theta} |q_0(\theta)| < \infty$. Moreover, from assumption H5, the function $\theta \mapsto q_0(\theta)$ is continuous and Theorem 6.5 leads to the result.

In the sequel, we use Z_t to denote $(X_{t-1}, \dots, X_{t-p})$, $t \in \mathbb{Z}$.

Lemma 6.4 *Let $t \in \mathbb{Z}$. Then for any realization z_t of Z_t, the distribution support of the random variable $X_{t|Z_t=z_t}$ has at least five points if $\sigma \neq 0$ and at least three, if $\sigma = 0$.*

Proof. The distribution of $X_{t|Z_t=z_t}$ is the same as the distribution of $C_{z_t} + \eta_t$ with

$$C_{z_t} = \sum_{j=1}^{p} \alpha_j \circ x_{t-j} + \varepsilon_t \sum_{j=1}^{p} \beta_j \circ x_{t-j}.$$

By H3 and using the fact that η_t are C_{z_t} independent, the result follows.
Lemma 6.5 will also be required for Theorem 6.4 demonstration.

Lemma 6.5 *Let $t \in \mathbb{Z}$. We have :*

1. *If $\sum_{j=1}^{p} \gamma_j X_{t-j} = \gamma$ then $\gamma = \gamma_j = 0, \forall j \in \{1, \dots, p\}$.*
2. *If we suppose $\sigma \neq 0$ and $\left(\sum_{j=1}^{p} s_j X_{t-j} \right) \left(\sum_{j=1}^{p} u_j X_{t-j} \right) + \sum_{j=1}^{p} \gamma_j |X_{t-j}| = \gamma$, then either $s_j = \gamma_j = \gamma = 0, \forall j \in \{1, \dots, p\}$ or $u_j = \gamma_j = \gamma = 0, \forall j \in \{1, \dots, p\}$.*
3. *If we suppose $\sigma = 0$ and $\sum_{j=1}^{p} \gamma_j |X_{t-j}| = \gamma$, then $\gamma_j = \gamma = 0$, $j = 1, \dots, p$.*

Proof.

1. suppose $m = \min\{j \in \{1, \dots, p\} : \gamma_j \neq 0\}$ exists. Then, X_{t-m} is measurable with respect to $\mathcal{F}_{t-m-1}$. This is in contradiction of Lemma 6.4. Hence, we deduce that $\gamma_j = 0$, $\forall j \in \{1, \dots, p\}$ from which it follows that $\gamma = 0$.

2. Suppose that $m = \min\{j \in \{1, \ldots, p\} : |s_j| + |u_j| + |\gamma_j| \neq 0\}$ exists. Note that if m does not exist, the result is obviously true.

   Suppose first that $m \leq p - 1$. Let

$$F(Z_{t-m}) = \sum_{j=m+1}^{p} (u_m s_j + s_m u_j) X_{t-j}$$

   and

$$G(Z_{t-m}) = \gamma - \left(\sum_{j=m+1}^{p} s_j X_{t-j} \right) \left(\sum_{j=m+1}^{p} u_j X_{t-j} \right) - \sum_{j=m+1}^{p} \gamma_j |X_{t-j}|$$

   We have

$$s_m u_m X_{t-m}^2 + F(Z_{t-m}) X_{t-m} + \gamma_m |X_{t-m}| = G(Z_{t-m}).$$

   Using Lemma 6.4, we see that for any realization z_{t-m} of Z_{t-m}, there exist five solutions to the equation with unknown x :

$$s_m u_m x^2 + F(z_{t-m}) x + \gamma_m |x| = G(z_{t-m}).$$

   Consequently, $s_m u_m = 0$, $G(z_{t-m}) = 0$ and $|F(z_{t-m})| = |\gamma_m|$.

   Without loss of generality, suppose $s_m = 0$. Then, the random variable $F(Z_{t-m}) = \sum_{j=m+1}^{p} u_m s_j X_{t-j}$ can take only two values almost surely : $\pm\gamma_m$. If $u_m = 0$, then $\gamma_m = 0$ and this is in contradiction of the assumption that m exists. Hence, $u_m \neq 0$. suppose $r = \min\{j : m+1 \leq j \leq p, s_j \neq 0\}$ exists. As

$$\sum_{j=r}^{p} s_j X_{t-j} \in \{\gamma_m / u_m, -\gamma_m / u_m\}$$

   we conclude that for any realization z_{t-r} of the random vector Z_{t-r}, the distribution support of the conditional law $X_{t-r|Z_{t-r}=z_{t-r}}$ has two points. This is in contradiction of Lemma 6.4. Hence, $s_j = 0, \forall j \in \{1, \ldots, p\}$. Hence equality $G(Z_{t-m}) = 0$ a.s leads to

$$\sum_{j=m+1}^{p} \gamma_j |X_{t-j}| = \gamma \quad \text{a.s.}$$

   If $q = \inf\{j \in \{m+1, \ldots, p\} : \gamma_j \neq 0\}$ exists, the distribution support of the conditional law $X_{t-q|Z_{t-q}=z_{t-q}}$ contains only two values. This is in contradiction of Lemma 6.4. So, $\gamma_j = 0, \forall j \geq 1$. Finally $\gamma = 0$ and the result follows.

   In the case where $m = p$, we have $s_p u_p X_{t-p}^2 + \gamma_p |X_{t-p}| = \gamma$. By Lemma 6.4, we conclude that necessarily $s_p u_p = \gamma_p = \gamma = 0$.

3. Suppose $m = \min\{j \in \{1, \ldots, p\} : \gamma_j \neq 0\}$ exists. Then, $|X_{t-m}|$ is measurable with respect to $\mathcal{F}_{t-m-1}$. Hence for each z_{t-m} the distribution support of the conditional law of $X_{t-m}/Z_{t-m} = z_{t-m}$ contains at most two points. This is in contradiction of Lemma (6.4) and the result follows.

Lemma 6.6 *If, $m_0(\theta) = m_0(\theta_0)$ and $V_0(\theta) = V_0(\theta_0)$ are satisfied, then $\theta = \theta_0$.*

Proof. Let us suppose $m_0(\theta) = m_0(\theta_0)$. Applying the first point of Lemma 6.5 with $\gamma_j = b_j(\theta) - b_j(\theta_0)$, $1 \le j \le p$ and $\gamma = \mu(\theta_0) - \mu(\theta)$, we obtain $b_j(\theta) = b_j(\theta_0)$, $j = 1, \ldots, p$ and $\mu(\theta) = \mu(\theta_0)$. More, suppose $V_0(\theta) = V_0(\theta_0)$. Two cases need to be considered.

- Firstly, assume that $\sigma^2 = \operatorname{Var} \varepsilon_t \neq 0$. We apply result 2. of Lemma 6.5 setting for $j \in \{1, \ldots, p\}$:

$$s_j = \sigma(c_j(\theta) - c_j(\theta_0)), \quad u_j = \sigma(c_j(\theta) + c_j(\theta_0)), \quad \gamma_j = w_j(\theta) - w_j(\theta_0),$$

  and $\gamma = \nu(\theta_0) - \nu(\theta)$. Then, it is easily seen that $w_j(\theta) = w_j(\theta_0)$, $j = 1, \ldots, p$, and $\nu(\theta) = \nu(\theta_0)$. Moreover, we have either $c_j(\theta) = c_j(\theta_0), \forall j \in \{1, \ldots, p\}$, either $c_j(\theta) = -c_j(\theta_0), \quad \forall j \in \{1, \ldots, p\}$. From the fact that there exists $j_0 \in \{1, \ldots, p\}$ such that the function c_{j_0} is positive and $c_{j_0}(\theta_0) > 0$, we can only have $c_j(\theta) = c_j(\theta_0)$, $j = 1, \ldots, p$.
- Secondly, assume that $\sigma = 0$. By the third point of Lemma 6.5 applied with $\gamma = \nu(\theta_0) - \nu(\theta)$ and $\gamma_j = w_j(\theta) - w_j(\theta_0)$, $1 \le j \le p$, we get $\nu(\theta) = \nu(\theta_0)$ and $w_j(\theta) = w_j(\theta_0)$, $j = 1, \ldots, p$.

The final conclusion, $\theta = \theta_0$, follows from H5. Now we can give Theorem 6.4 demonstration. In fact, it is done in a very classical way. By Lemma 6.3, we have :

$$\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| \to_{T \to \infty} 0 \quad \text{a.s.}$$

Lemma 6.6 can be used to show that

$$Q(\theta_0) < Q(\theta), \quad \forall \theta \in \Theta \setminus \{\theta_0\}$$

(see for example the proof of proposition 2.1 in [15]).
From these last two properties, a classical compactness argument leads to the strong consistency of $\hat{\theta}_T$ (see for example [23, Theorem 2.2.1, p. 19]).

6.6.6 Proof of Theorem 6.1

Let $t \in \mathbb{Z}$. In Section 6.4.1, $q_t(\theta)$ has been defined as :

$$q_t(\theta) = \frac{(X_t - m_t(\theta))^2}{V_t(\theta)} + \ln V_t(\theta).$$

So the first and second derivatives are :

$$\nabla q_t(\theta) = \frac{\nabla V_t(\theta)}{V_t(\theta)}\left(1 - \frac{(X_t - m_t(\theta))^2}{V_t(\theta)}\right) - 2\frac{(X_t - m_t(\theta))\nabla m_t(\theta)}{V_t(\theta)} \tag{6.13}$$

$$\begin{aligned}\nabla^2 q_t(\theta) = \frac{1}{V_t(\theta)^2}\Big[&\nabla V_t(\theta)\nabla V_t(\theta)'\left(2\frac{(X_t - m_t(\theta))^2}{V_t(\theta)} - 1\right)\\ &+ \nabla^2 V_t(\theta)\left(1 - \frac{(X_t - m_t(\theta))^2}{V_t(\theta)}\right) + 2(X_t - m_t(\theta))\nabla m_t(\theta)\nabla V_t(\theta)'\\ &+ 2V_t(\theta)\nabla m_t(\theta)\nabla m_t(\theta)' - 2V_t(\theta)(X_t - m_t(\theta))\nabla^2 m_t(\theta)\\ &+ 2\left(X_t - m_t(\theta)\right)\nabla V_t(\theta)\nabla m_t(\theta)'\Big]\end{aligned} \tag{6.14}$$

Lemmas 6.7 to 6.9 give important properties of $\nabla q_t(\theta)$ and $\nabla^2 q_t(\theta)$. They are required to prove Theorem 6.1

Lemma 6.7 *For all $\theta \in \Theta$, the sequences $\{\nabla q_t(\theta)\}_t$ and $\{\nabla^2 q_t(\theta)\}_t$ are ergodic and stationary.*

Proof. We use the same argument than the one we gave in Lemma 6.3 proof to show that the sequence $(q_t(\theta))_t$ is stationary and ergodic.
From now on, $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$ or the matrix norm associated with, as required.

Lemma 6.8 *We have :*

$$\mathbb{E}\left\|\nabla q_0(\theta_0)\right\|^2 < \infty \text{ and } \mathbb{E}\sup_{\theta\in\Theta}\left\|\nabla^2 q_0(\theta)\right\| < \infty.$$

Proof. Recall that if P is a polynomial of degree q defined on $\mathbb{R}^p$, then there exist non-negative constants $d_0, \ldots, d_p$ such that :

$$\left|P\left(X_{-1}, \ldots, X_{-p}\right)\right| \le d_0 + \sum_{j=1}^{p} d_j \left|X_{-j}\right|^q, \quad \text{a.s.}$$

Proof of the first assertion : $\mathbb{E}\left\|\nabla q_0(\theta_0)^2\right\| < \infty$.

– We first observe that the ratio :
$$\frac{(X_0 - m_0(\theta_0))\nabla m_0(\theta_0)}{V_0(\theta_0)}$$
is square integrable. Indeed, as
$$\mathbb{E}\left(\frac{(X_0 - m_0(\theta_0))^2\left\|\nabla m_0(\theta_0)\right\|^2}{V_0(\theta_0)^2}/\mathcal{F}_{-1}\right) = \frac{\left\|\nabla m_0(\theta_0)\right\|^2}{V_0(\theta_0)} \le \frac{\left\|\nabla m_0(\theta_0)\right\|^2}{h}$$
there exist positive constants $d_0, \ldots, d_p$ such that
$$\mathbb{E}\left(\frac{(X_0 - m_0(\theta_0))^2\left\|\nabla m_0(\theta_0)\right\|^2}{V_0(\theta_0)^2}/\mathcal{F}_{-1}\right) \le d_0 + \sum_{j=1}^{p} d_j X_{-j}^2$$

and the integrability follows from $\mathbb{E}X_0^4 < \infty$.

– Now, we show that $\frac{\nabla V_0(\theta_0)}{V_0(\theta_0)}\left(1 - \frac{(X_0 - m_0(\theta_0))^2}{V_0(\theta_0)}\right)$ is square integrable.
As $V_0(\theta_0) \geq h$ a.s. and $\mathbb{E}X_0^4 < \infty$, it is easily seen that the ratio $\frac{\nabla V_0(\theta_0)}{V_0(\theta_0)}$ is square integrable. Then it is enough to show that the variable $V_0(\theta_0)^{-2}\nabla V_0(\theta_0)(X_0 - m_0(\theta_0))^2$ is square integrable. Let

$$C_0(\theta_0) = (X_0 - m_0(\theta_0))^2.$$

As $\nabla V_0(\theta_0)$ is square integrable and measurable with respect to $\mathcal{F}_{-1}$, it is sufficient to show that the random variable $V_0^{-4}(\theta_0)\mathbb{E}\left(C_0^2(\theta_0)/\mathcal{F}_{-1}\right)$ is bounded.
We notice that :

$$\begin{aligned}
&\mathbb{E}\left(C_0(\theta_0)^2/X_{-1} = x_{-1}, \ldots, X_{-p} = x_{-p}\right) \\
&= \left\|\sum_{j=1}^p b_j(\theta_0) \circ x_{-j} + \xi_0 \sum_{j=1}^p c_j(\theta_0) \circ x_{-j} - \sum_{j=1}^p w_j(\theta_0)\left|x_{-j}\right| + \eta_t\right\|_4^4 \\
&\leq \left[\sum_{j=1}^p \left\|Y_{0,0}^{(j)}\right\|_4 \left|x_{-j}\right| + \|\xi_0\|_4 \left(\sum_{j=1}^p \left\|Z_{0,0}^{(j)}\right\|_4 \left|x_{-j}\right|\right) + \sum_{j=1}^p w_j(\theta_0)\left|x_{-j}\right| + \|\eta_0\|_4\right]^4.
\end{aligned}$$

We deduce that there exist constants $d_0, \ldots, d_p$ such that

$$\mathbb{E}\left(C_0(\theta_0)^2/\mathcal{F}_{-1}\right) \leq d_0 + \sum_{j=1}^p d_j \left|X_{-j}\right|^4.$$

Since $V_0(\theta_0) \geq w_j(\theta_0)\left|X_{-j}\right| \wedge h$ for $j = 1, \ldots, p$, it follows that :

$$V_0(\theta_0)^{-4}\mathbb{E}\left(C_0(\theta_0)^2/\mathcal{F}_{-1}\right) \leq \frac{d_0}{h^4} + \sum_{j=1}^p \frac{d_j}{w_j(\theta_0)^4}.$$

We have shown that $V_0(\theta_0)^{-4}\mathbb{E}\left(C_0(\theta_0)^2/\mathcal{F}_{-1}\right)$ is bounded.

Proof of the second assertion : $\mathbb{E}\sup_{\theta\in\Theta}\left\|\nabla^2 q_0(\theta)\right\| < \infty$. We start the demonstration by showing that

$$\mathbb{E}\sup_{\theta\in\Theta}\left(\|\nabla V_0(\theta)\|^2 V_0(\theta)^{-3}[X_0 - m_0(\theta)]^2\right) < \infty. \tag{6.15}$$

For this, we have :

$$\begin{aligned}
\mathbb{E}\left(\sup_{\theta\in\Theta} V_0^{-3}(\theta)[X_0 - m_0(\theta)]^2/\mathcal{F}_{-1}\right) &\leq \frac{2}{h\inf_{\theta\in\Theta} V_0^2(\theta)}\mathbb{E}\left(X_0^2 + \left(\sum_{j=1}^p \|b_j\|_\infty X_{-j} + \|\mu\|_\infty\right)^2/\mathcal{F}_{-1}\right) \\
&\leq \frac{2\left(m_0(\theta_0)^2 + V_0(\theta_0) + \left(\sum_{j=1}^p \|b_j\|_\infty X_{-j} + \|\mu\|_\infty\right)^2\right)}{h\inf_{\theta\in\Theta} V_0^2(\theta)}.
\end{aligned}$$

Hence, there exist non-negative constants $d_0, \dots, d_p$ such that

$$\mathbb{E}\left(\sup_{\theta\in\theta} V_0^{-3}(\theta)[X_0 - m_0(\theta)]^2/\mathcal{F}_{-1}\right) \leq \frac{1}{\inf_{\theta\in\Theta} V_0^2(\theta)}\left(d + \sum_{j=1}^{p} d_j X_{-j}^2\right).$$

As $\frac{X_{-j}^2}{\inf_{\theta\in\theta} V_t^2(\theta)} \leq \frac{1}{\inf_{j\in\theta} w_j^2(\theta)}$, according to hypothesis H8, we conclude that there exists a constant $M > 0$ such that

$$\mathbb{E}\left(\sup_{\theta\in\Theta} V_0^{-3}(\theta)[X_0 - m_0(\theta)]^2/\mathcal{F}_{-1}\right) < M. \tag{6.16}$$

We note that using H8 only for $i = 1, \dots, d$,

$$\begin{aligned}\left\|\frac{\partial V_0}{\partial \theta_i}\right\|_\infty &\leq 2\sigma^2\left(\sum_{j=1}^{p}\|c_j\|_\infty |X_{-j}|\right)\left(\sum_{j=1}^{p}\left\|\frac{\partial c_j}{\partial \theta_i}\right\|_\infty |X_{-j}|\right) \\ &+\sum_{j=1}^{p}\left\|\frac{\partial w_j}{\partial \theta_i}\right\|_\infty |X_{-j}| + \|\nu\|_\infty .\end{aligned}$$

Whence, we conclude that if $\mathbb{E}X_0^4 < \infty$ then $\mathbb{E}\sup_{\theta\in\Theta}\left\|\nabla V_0(\theta)^2\right\| < \infty$.
From (6.16) follows (7.20).
We notice that the other terms of (6.14) are uniformly bounded by polynomials of the fourth degree in $|X_{-1}|, \dots, |X_{-p}|$. This completes the proof.

Lemma 6.9 *The entries of the column vectors of the differential of the function $\theta \mapsto (m_0(\theta), V_0(\theta))$ evaluated at θ_0 are linearly independent random variables.*

Proof. The proof is done in three steps.

Step 1. Suppose there exist constants $\lambda_1, \dots, \lambda_d$ such that

$$\sum_{i=1}^{d}\lambda_i\frac{\partial m_0}{\partial \theta_i}(\theta_0) = 0 \quad \text{a.s} \quad \text{or} \quad \sum_{i=1}^{d}\lambda_i\frac{\partial V_0}{\partial \theta_i}(\theta_0) = 0 \quad \text{a.s.}$$

Since $m_0(\theta_0) = \mu(\theta_0) + \sum_{j=1}^{p} b_j(\theta_0)X_{-j}$, the partial derivatives at θ_0 are :

$$\frac{\partial m_0}{\partial \theta_i}(\theta_0) = \frac{\partial \mu}{\partial \theta_i}(\theta_0) + \sum_{j=1}^{p}\frac{\partial b_j}{\partial \theta_i}(\theta_0)X_{-j}, \quad i = 1, \dots, d.$$

Then we have :

$$\sum_{i=1}^{d}\lambda_i\frac{\partial m_0}{\partial \theta_i}(\theta_0) = \sum_{i=1}^{d}\lambda_i\frac{\partial \mu}{\partial \theta_i}(\theta_0) + \sum_{j=1}^{p}\sum_{i=1}^{d}\lambda_i\frac{\partial b_j}{\partial \theta_i}(\theta_0)X_{-j} = 0 \quad \text{a.s.}$$

By the first result of Lemma 6.5, we get :

$$\sum_{i=1}^{d}\lambda_i\frac{\partial \mu}{\partial \theta_i}(\theta_0) = \sum_{i=1}^{d}\lambda_i\frac{\partial b_j}{\partial \theta_i}(\theta_0) = 0 \quad \text{a.s,} \quad j = 1, \dots, p.$$

Step 2. Since $V_0(\theta_0) = \left(\sigma\sum_{j=1}^{p} c_j(\theta_0)X_{-j}\right)^2 + \sum_{j=1}^{p} w_j(\theta_0)\,|X_{-j}| + \nu(\theta_0)$, the partial derivatives at θ_0 are :

$$\frac{\partial V_0}{\partial \theta_i}(\theta_0) = 2\sigma^2 \left(\sum_{j=1}^{p} c_j(\theta_0)X_{-j}\right)\left(\sum_{j=1}^{p} \frac{\partial b_j}{\partial \theta_i}(\theta_0)X_{-j}\right) + \sum_{j=1}^{p} \frac{\partial w_j}{\partial \theta_i}(\theta_0)\,|X_{-j}| + \frac{\partial \nu}{\partial \theta_i}(\theta_0),$$

for $i = 1, \ldots, d$. So, we have :

$$\sum_{i=1}^{d} \lambda_i \frac{\partial V_0}{\partial \theta_i}(\theta_0) = 2\sigma^2 \left(\sum_{j=1}^{p} c_j(\theta_0)X_{-j}\right)\left(\sum_{j=1}^{p}\sum_{i=1}^{d} \lambda_i \frac{\partial c_j}{\partial \theta_i}(\theta_0)X_{-j}\right) + \sum_{j=1}^{p}\sum_{i=1}^{d} \lambda_i \frac{\partial w_j}{\partial \theta_i}(\theta_0)\,|X_{-j}| + \sum_{i=1}^{d} \lambda_i \frac{\partial \nu}{\partial \theta_i}(\theta_0) = 0.$$

The last equation can be written as :

$$2\sigma^2 \left(\sum_{j=1}^{p} c_j(\theta_0)X_{-j}\right)\left(\sum_{j=1}^{p}\sum_{i=1}^{d} \lambda_i \frac{\partial c_j}{\partial \theta_i}(\theta_0)X_{-j}\right) + \sum_{j=1}^{p}\sum_{i=1}^{d} \lambda_i \frac{\partial w_j}{\partial \theta_i}(\theta_0)\,|X_{-j}| = -\sum_{i=1}^{d} \lambda_i \frac{\partial \nu}{\partial \theta_i}(\theta_0).$$

Using the second result of Lemma 6.5 and the assumption that $c_{j_0}(\theta_0) = \beta_{j_0} > 0$, we get

$$2\sigma^2 \sum_{i=1}^{d} \lambda_i \frac{\partial c_j}{\partial \theta_i}(\theta_0) = \sum_{i=1}^{d} \lambda_i \frac{\partial w_j}{\partial \theta_i}(\theta_0) = \sum_{i=1}^{d} \lambda_i \frac{\partial \nu}{\partial \theta_i}(\theta_0) = 0, \quad j = 1, \ldots, p.$$

Step 3. By the last two steps and hypothesis H8, we deduce that $\lambda_1 = \cdots = \lambda_d = 0$ which shows that the entries of the vectors of the differential of the bifold function $(m_0(\theta), V_0(\theta))$ evaluated at θ_0 are linearly independent random variables.

We are now ready to prove Theorem 6.1.

Proof of Theorem 6.1. The technique for the proof of this theorem is very classical, we follow the proof given in [23, Theorem 2.2.1, p. 19]. Since $\theta \in \Theta^\circ$, using a Taylor expansion, we get :

$$0 = \nabla Q_T(\hat{\theta}_T) = \nabla Q_T(\theta_0) + \widetilde{M}_T \cdot (\hat{\theta}_T - \theta_0)$$

where $\widetilde{M}_T$ is the matrix of the second order derivatives, that is :

$$\widetilde{M}_T(i,j) = \frac{\partial^2 Q_T}{\partial \theta_i \partial \theta_j}(\gamma_i), \quad 1 \leqslant i,\, j \leqslant d.$$

with $\left\|\hat{\theta}_T - \gamma_i\right\| \leq \left\|\hat{\theta}_T - \theta_0\right\|, \quad i = 1, \dots, d$. Hence,

$$\sqrt{T}Q_T(\theta_0) = \sqrt{T}\widetilde{M}_T \cdot (\hat{\theta}_T - \theta_0)$$

By Lemma 6.7, $\left(\nabla^2 q_t(\theta_0)\right)_{t\in\mathbb{Z}}$ is an ergodic stationary sequence. By hypothesis H8, its values are in $\mathcal{C}(\Theta, \mathbb{R}^d \times \mathbb{R}^d)$. According to Lemma 6.8, $\sup_{\theta\in\Theta}\left\|\nabla^2 q_0(\theta)\right\|$ is integrable. Then, we can apply Theorem 6.5 and since $\hat{\theta}_T \to_{T\to\infty} \theta_0$ a.s, we conclude that $\widetilde{M}_T \to_{T\to\infty} F_0 = \mathbb{E}\nabla^2 q_t(\theta_0)$ a.s. More, F_0 is non-singular. Indeed :

$$F_0 = \mathbb{E}\left(V_0(\theta_0)^{-2}\left\{\nabla V_0(\theta_0)\nabla V_0(\theta_0)' + 2V_0(\theta_0)\nabla m_0(\theta_0)\nabla m_0(\theta_0)'\right\}\right),$$

and by Lemma 6.9, this matrix is positive-definite. More :

$$\sqrt{T}Q_T(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla q_t(\theta_0) \quad \text{and} \quad \mathbb{E}\left(\nabla q_t(\theta_0)/\mathcal{F}_{-1}\right) = 0.$$

Since by Lemma 6.8, $\mathbb{E}\left\|\nabla q_0(\theta_0)\right\|^2 < \infty$, the sequence $\left(\nabla q_t(\theta_0)\right)_t$ is an ergodic stationary $\mathcal{F}_t$-martingale difference sequence of finite variance. Then by [1, Theorem 23.1, p. 206], we have : $\sqrt{T}Q_T(\theta_0) \to_{T\to\infty} \mathcal{N}(0, G_0)$ in distribution, with $G_0 = \mathbb{E}\left(\nabla q_0(\theta_0)\nabla q_0(\theta_0)'\right)$. Consequently, we get :

$$\sqrt{T}(\hat{\theta_T} - \theta_0) \to_{T\to\infty} \mathcal{N}(F_0^{-1}G_0F_0^{-1}).$$

The expression of G_0 follows from straightforward computations using the expression (6.13).

Bibliographie

[1] Billingsley, P. (1999) *Convergence of probability measures.* second ed. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley and Sons Inc., New York. A Wiley-Interscience Publication.

[2] Dion, J.-P., Gauthier, G., Latour, A. (1995) *Branching processes with immigration and integer-valued time series.* Serdica Math. J. 21, 123-136.

[3] Doukhan, P., Latour, A., Oraichi, D. (2006) *A simple integer-valued bilinear times series models.* Advances in Applied Probability 38, 1-20.

[4] Doukhan, P., Truquet, L. (2007) *Weakly dependent random fields with infinite memory.* ALEA : Latin American Journal of Probability and Mathematical Statistics 2, 111-132.

[5] Drost, F. C., Van Den Akker, R., Werker, B. J. (2007) *Note on integervalued bilinear time series models.* Statistics and Probability Letters.

[6] Du, J.-G., Li, Y. (1991) *The integer-valued autoregressive (INAR(p)).* model. J. Times Ser. Anal. 12, 129-142.

[7] Enciso-Mora, V., Neal, P., Subba Rao, T. (2006) *Efficient order selection algorithms for integer valeud ARMA processes.* Technical report 2006.415. The University of Manchester School of Mathematics, The University of Manchester, Manchesterm M13 9PL, UK.

[8] Ferland, R., Latour, A., Oraichi, D. (2006) *Integer-valued GARCH process.* Journal of Time Series Analysis 27, 923-942.

[9] Fokianos, K. (2004) *Integer autoregressive and moving average models.* Ann. I.S.U.P. 48, 53-58.

[10] Fokianos, K., Fried, R. (2008) *Intervention effects in INGARCH processes.* Working paper.

[11] Freeland, R., Mccab, B. (2004) *Forecasting discrete valued low count time series.* International Journal of Forecasting 20, 427-434.

[12] Gather, U., Schettlinger, K., Fried, R. (2006) *Online signal extraction by robust linear regression.*Computational Statistics 21, 33-51.

[13] Heinen, A. (2001) *Modeling Time Series Count Data : The Autoregressive Conditional Poisson Model.* PhD thesis. Department of Economics, University of California San Diego, California.

[14] Hellström, J. (2001) *Unit root testing in integer-valued ar(1) models.* Economics Letters 70, 9-14.

[15] Jeantheau, T. (1998) *Strong consistency of estimators for multivariate ARCH models.* Econometric Theory 14, 70-86.

[16] Kedem, B., Fokianos, K. (2002) *Regression models for binary time series.* In Modeling uncertainty. vol. 46 of Internat. Ser. Oper. Res. Management Sci. Kluwer Academic Publisher, Boston, MA. pp. 185-199.

[17] Kedem, B., Fokianos, K. (2002) *Regression Models for Time Series Analysis.* Wiley series in probability and statistics. Wiley-Interscience, Hoboken, New Jersey.

[18] Latour, A. (1998) *Existence and stochastic structure of a non-negative integervalued autoregressive process.* J. Time Ser. Anal. 19, 439-455.

[19] Macdonald, I. L., Zucchini, W. (1997) *Hidden Markov and other models for discrete-valued time series.* Vol. 70 of Monographs on Statistics and Applied Probability. Chapman and Hall, London.

[20] Mccab, B., Martin, G. (2005) *Bayesian predictions of low count time series.* International Journal of Forecasting 22, 315-330.

[21] Neal, P., Subba Rao, T. (2007) *Mcmc for integer-valued ARMA processes.* Journal of Time Series Analysis 28, 92-110.

[22] Steutel, F., Van Harn, K. (1979) *Discrete analogues of selfdecomposability and stability.* Ann. Prob. 7, 893-899.

[23] Straumann, D. (2005) *Estimation in conditionally heteroscedastic time series models.* Vol. 181 of Lecture Notes in Statistics. Springer-Verlag, Berlin.

Chapitre 7

A new smoothed QMLE for AR processes with LARCH errors

Abstract

We introduce a smoothed version of the quasi maximum likelihood estimator (QMLE) in order to fit heteroschedastic time series with possibly vanishing conditional variance. We apply this procedure to a finite-order autoregressive process with linear ARCH errors. We prove both the almost sure consistency and the asymptotic normality of our estimator. This estimator is more robust that QMLE with the same type of assumptions. A numerical study confirms the qualities of our procedure.

7.1 Introduction

In order to study the behaviour of financial time series such as asset returns or exchange rates, a considerable work has been done to study ARCH models introduced by Engle (1982) [8]. From empirical observations of time series, Black (1976) [3] called the leverage effect a tendency for the conditional variance to be negatively correlated with the past returns. Another property is a slow decay of the autocorrelations of the squares called long memory, see Doukhan *et al.* [6]. LARCH(∞) (Linear ARCH(∞)) models introduced in Giraitis *et al.* (2000) [14] and (2004) [15] take in account these two properties ; they are defined from an equation :

$$X_t = \xi_t \Big(a_0 + \sum_{j \geq 1} a_j X_{t-j} \Big) \tag{7.1}$$

for an independent and identically distributed (i.i.d) sequence (ξ_t) with $\mathbb{E}\xi_0 = 0$ and $\mathbb{E}\xi_0^2 = 1$. Long memory properties of the model are addressed by Giraitis *et al.* (2000) [14] whereas the leverage property is studied in Giraitis *et al.* (2004) [15]. The model (7.1) specializes to the asymmetric

ARCH model of Engle (1990)[9]. The conditional variance of models (7.1) writes as the square of a linear combination of the past values :

$$V_t = \Big(a_0 + \sum_{j\geq 1} a_j X_{t-j}\Big)^2$$

A short memory version of (7.1) is LARCH(p), here $a_j = 0$ for $j > p$ hence :

$$X_t = \xi_t \Big(a_0 + \sum_{j=1}^{p} a_j X_{t-j}\Big) \tag{7.2}$$

Then the model (7.2) is a special case of a more general model introduced in Sentena (1995) [20]; here the conditional variance writes as a quadratic form :

$$V_t = a_0 + \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=1}^{p}\sum_{k=1}^{p} b_{j,k} X_{t-j} X_{t-k} \tag{7.3}$$

If this quadratic form is nonnegative this is possible to exhibit assumptions ensuring $V_t \geqslant 0$; a solution (X_t) with this conditional variance writes $X_t = \xi_t V_t^{1/2}$ for iid inputs ξ_t. Giraitis *et al.* (2000) [14] prove a necessary and sufficient condition for the existence and the uniqueness of a square integrable and strictly stationary solution of the equation (7.1). Sufficient conditions for the existence of higher moments is also provided, moreover Giraitis *et al.* (2004) [15] explicit sufficient conditions for the leverage property.

The model (7.1) is generalized in Giraitis and Surgailis (2002) [16] to a bilinear model which exhibits long memory both in conditional mean and in conditional variance :

$$X_t = \alpha + \sum_{j\geq 1} \alpha_j X_{t-j} + \xi_t \Big(a_0 + \sum_{j\geq 1} a_j X_{t-j}\Big) \tag{7.4}$$

In the short memory case, Francq *et al.* (2008) [11] study existence and uniqueness of a strictly stationary solution of equation (7.2) (not necessarily square integrable).

A main statistical problem is to estimate the parameter $\theta = (a_0, \ldots, a_p)$ of the model (7.2). In a recent work, Beran and Schützner (2008) [2] consider the estimation of the parameters C and d when the coefficients in equation (7.1) have the form $a_j = Cj^d$, both in the short and long memory cases. They use a modified conditional maximum likelihood estimator and the same approach will be used in this paper. On the other hand another recent work by Francq and Zakoïan [12] shows the consistency and asymptotic normality of a Weighted Least Squares estimation for model (7.2). The principle of their method is to apply the least square prodedure to the square of the process.

A classical estimation procedure is the Gaussian Quasi Maximum Likelihood Estimation (QMLE). Under conditions, the QMLE is shown to be consistent and asymptotically normal. But a crucial condition in its application is the existence of a real number $h > 0$ such that $V_0(\theta) \geq h$ a.s. For the model (7.2), the conditional variance V_0 is, in general, not bounded away from 0 and the quasi likelihood becomes numerically intractable. Because the QMLE cannot be used for the model (7.2), we propose a smoothed version of the QMLE which is more robust then the classical QMLE and applies with the same kind of assumptions. We apply this procedure to an AR process with LARCH errors.

The paper is organized as follows. Section 2 recalls the properties of the model (7.2). The next Section 3 introduces our model and mand motivates the introduction of our smoothed QMLE. Section 4 addresses its asymptotic properties for our model. In Section 5, we discuss the behaviour of its asymptotic variance when the smoothing parameter tends to 0. Section 6 is dedicated to a numerical illustration. The proofs are postponed to the last section of the paper.

7.2 Some general results about LARCH models.

The first results about existence of LARCH models were given in the general case of equation (7.1). The condition $\sum_{j=1}^{\infty} a_j^2 < 1$ is necessary and sufficient for the existence of a square integrable and nonanticipative solution (see Theorem 2.1 in Giraitis *et al.* (2004) [15]). Those authors prove that the unique solution of equation (7.1) is defined from the Volterra expansion :

$$X_t = a_0\xi_t\Big(1+\sum_{k\geq 1}\sum_{j_1,\ldots,j_k\geq 1} a_{j_1}\cdots a_{j_k}\xi_{t-j_1}\cdots\xi_{t-(j_1+\cdots+j_k)}\Big) \tag{7.5}$$

Those authors also give a sufficient condition for the existence of the fourth moments in the general case of model (7.1) :

$$\mu_4\sum_{j=1}^{\infty} a_j^4 + 4\,|\mu_3|\sum_{j=1}^{\infty}|a_j|^3 + 6\sum_{j=1}^{\infty} a_j^2 < 1 \tag{7.6}$$

where $\mu_i = \mathbb{E}\xi_0^i$ for $1 \leqslant i \leqslant 4$ (here $\mu_2 = 1$). Mention that

$$\mu_4^{1/4}\sum_{j=1}^{\infty}|a_j| < 1 \tag{7.7}$$

ensures the existence of the fourth moment for the solution (7.5) (see Doukhan *et al.* (2006) [7]), hence the condition (7.6) is perhaps not sharp. Although condition (7.6) is less restrictive with respect to the decay of the sequence $(a_j)_{j\geq 1}$, condition (7.7) can obviously be better, *e.g.* if $a_j = 0$, $j \geq 2$.

From now on we fix an integer $p \geq 1$ and we only consider eqn. (7.2). Then, existence and uniqueness of a strictly stationary solution of (7.2) holds under the less restrictive condition $\sum_{j=1}^{p} a_j^2 < 1$, is pointed in Francq *et al.* (2008) [11]. Denote

$$A_t = \begin{pmatrix} a_{1:p-1}\xi_t & a_p\xi_t \\ I_{p-1} & 0_{p-1} \end{pmatrix}, \qquad \text{where} \qquad a_{1:p-1} = (a_1,\ldots,a_{p-1})$$

and I_k is the $k \times k$ identity matrix. If $p=1$ then $A_t = a_1\xi_t$. Let $A = (A_t)_t$ and $\gamma(A)$ the top-Lyapunov exponent of the sequence A :

$$\gamma(A) = \lim_{t\to\infty} \frac{1}{t} \log \|A_t \cdots A_1\|$$

Theorem 3.1 in Francq *et al.* (2008) [11] asserts that equation (7.2) admits a strictly stationary solution if and only if

$$\gamma(A) < 0. \tag{7.8}$$

Under this condition, the strictly stationary solution is unique, nonanticipative and ergodic.
If $p=1$ condition (7.8) writes explicitly $|a_1| < e^{-\mathbb{E}\log|\xi_0|}$. As pointed in [11], if $\xi_0 \sim \mathcal{N}(0,1)$ this writes $|a_1| < 1.88736$. In comparison the condition for the existence of a second moment writes $|a_1| < 1$ and the one for the fourth moment $|a_1| < 0.7598\cdots$ (see section 7.3).

7.3 Model specification and smoothed QMLE

We consider for $p, q \in \mathbb{N}^*$ the model

$$\begin{aligned} Y_t &= b_{0,1}Y_{t-1} + \cdots + b_{0,q}Y_{t-q} + X_t, && (7.9)\\ X_t &= \xi_t\left(a_{0,0} + \sum_{j=1}^{p} a_{0,j}X_{t-j}\right), \qquad t \in \mathbb{Z} && (7.10)\end{aligned}$$

with ξ an i.i.d sequence such that $\mathbb{E}\xi_0 = 0$, $\mathbb{E}\xi_0^2 = 1$. By convention, $q = 0$ means that the process Y is a pure LARCH model given by (7.10) ; note that for $p = 0$ the model is an AR(q) process.
In the sequel, when we consider a solution of equation (7.9) or (7.10), it is always assumed that this solution is stationary, ergodic and non anticipative. We denote

$$\theta_0 = (b_{0,1}, \ldots, b_{0,q}, a_{0,0}, \ldots, a_{0,p}),$$

and for $\theta = (b_1, \ldots, b_q, a_0, \ldots, a_p) \in \mathbb{R}^{p+q+1}$ and $t \in \mathbb{Z}$:

$$\begin{aligned} m_t(\theta) &= \sum_{j=1}^{q} b_j Y_{t-j},\\ V_t(\theta) &= \sigma_t^2(\theta) = \Big(a_0 + \sum_{j=1}^{p} a_j\left(Y_{t-j} - m_{t-j}(\theta)\right)\Big)^2. \end{aligned}$$

Setting $\mathcal{F}_t = \sigma(Y_{t-1}, Y_{t-2}, \ldots)$ for $t \in \mathbb{Z}$, we have :

$$m_t(\theta_0) = \mathbb{E}\left(Y_t/\mathcal{F}_{t-1}\right), \qquad V_t(\theta_0) = \mathrm{Var}\left(Y_t/\mathcal{F}_{t-1}\right).$$

By stationarity, we can always suppose that the data $Y_n, Y_{n-1}, \ldots, Y_{-(p+q)+1}$ are available. Usually, the QMLE is defined by :

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta),$$

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\frac{(Y_t - m_t(\theta))^2}{V_t(\theta)} + \ln V_t(\theta). \tag{7.11}$$

Although we will not prove any result about the consistency or the inconsistency of the QMLE for the model (7.9), it seems very difficult to compute this estimator because of the intractable form of the function $\theta \mapsto Q_n(\theta)$ (see figure 7.1 for which $q = 0$, $p = 1$, $a_0 = 1$; the data are generated with $a_1 = 0.5$ and $\xi_0 \sim \mathcal{N}(0,1)$). The roughness of the function $\theta \mapsto Q_n(\theta)$ is due to the small values of the function $\theta \mapsto V_t(\theta)$; this gives infinite values for Q_n.

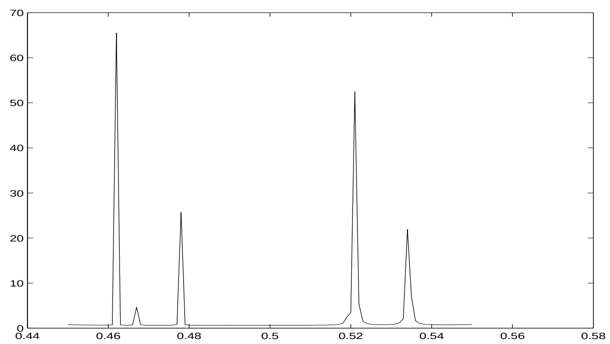


FIG. 7.1: $a_1 \mapsto Q_{500}(\theta)$.

In some cases, even the conditional variance $V_0(\theta_0)$ is not bounded away from zero as shows the following Lemma :

Lemma 7.1 *Suppose that for the model (7.10) the input ξ_0 admits a density with support $\mathbb{R}$ and $a_{0,0} \neq 0$. If $j_0 = \min\{j/a_{0,j} \neq 0\}$ exists, then the conditional variance $V_0(\theta_0)$ is unbounded away from zero.*

One also can note that even if the conditonal variance may be unbounded away from zero, it does not vanish if ξ_0 is atomless :

Lemma 7.2 *Suppose that for model (7.10), the law of ξ_0 is atomless, then* $\mathbb{P}\left(\sigma_0(\theta_0)=0\right)=0$.

We define here a new contrast working also will small values of the conditional variance. Suppose just for a moment (even this is not true) that (Y_t) is a ARCH process with a conditional variance bounded away from zero; if $\xi_0 \sim \mathcal{N}(0,1)$ then $-1/2(Q_n+\ln 2\pi)$ is well defined and is the exact conditional log-likelihood of the random vector $(Y_1,\dots,Y_n)$. For nonGaussian inputs, ξ_0, this is not true anymore but the function

$$Q:\theta\mapsto \mathbb{E}\left(\frac{(Y_0-m_0(\theta))^2}{V_0(\theta_0)}+\ln(V_0(\theta))\right) \tag{7.12}$$

is still a good contrast since θ_0 is the unique minimum of Q, provided the following identification condition holds :

$$(m_0(\theta),V_0(\theta))=(m_0(\theta_0),V_0(\theta_0))\Rightarrow\theta=\theta_0.$$

In fact, Q can be used as a contrast for the estimation of a parameter of the mean and/or the variance of a conditional law W_0/U_0, for some stationary ergodic process $\{(U_t,W_t)/t\in\mathbb{Z}\}$; in our case $U_t=\left(Y_{t-1},\dots,Y_{t-(p+q)}\right)$ and $W_t=Y_t$.
Set now $U_t=\left(Y_{t-1},\dots,Y_{t-(p+q)}\right)$ and $W_t=Y_t+\eta_t$, for $t\in\mathbb{Z}$, where $(\eta_t)_t$ denotes an i.i.d sequence, independent of the process (Y_t), with $\mathbb{E}\eta_0=0$ and $\mathrm{Var}\,(\eta_0)=h$ for some real number $h>0$. Then we have

$$\mathbb{E}\left(W_t/U_t\right)=m_t(\theta_0),\qquad \mathrm{Var}\;(W_t/U_t)=V_t(\theta_0)+h,$$

and the contrast Q becomes

$$Q_h(\theta)=\mathbb{E}\left(\frac{(W_0-m_0(\theta))^2}{V_0(\theta)+h}+\ln(V_0(\theta)+h)\right).$$

We obtain from independence,

$$Q_h(\theta)=\mathbb{E}\left(\frac{(Y_0-m_0(\theta))^2+h}{V_0(\theta)+h}+\ln(V_0(\theta)+h)\right).$$

The number $h>0$ avoids the problem of small possible values for the variance in (7.12) : it will be called the *smoothing parameter*. If the data $Y_n,Y_{n-1},\dots,Y_{-(p+q)+1}$ are available, we define the following estimator :

$$\hat{\theta}_{n,h} = \arg\min_{\theta \in \Theta} Q_{n,h}(\theta), \tag{7.13}$$

$$Q_{n,h}(\theta) = \frac{1}{n}\sum_{t=1}^{n} q_{t,h}(\theta), \tag{7.14}$$

$$q_{t,h}(\theta) = \frac{(Y_t - m_t(\theta))^2 + h}{V_t(\theta) + h} + \ln\left(V_t(\theta) + h\right). \tag{7.15}$$

Observe that $\hat{\theta}_{n,0}$ is the classical QMLE. For $h > 0$ and $n \in \mathbb{N}^*$, $Q_{n,h}$ has a more tractable expression that $Q_{n,0}$. The asymptotic properties of the estimator $\hat{\theta}_{n,h}$, called *smoothed QMLE*, will be derived below.

7.4 Asymptotics of smoothed QMLE for AR-LARCH models

QMLE is very popular for conditionally heteroscedastic time series. Its asymptotic properties were first established by Weiss (1986) [22] for ARCH models. General results for the consistency of this method are proved in Jeantheau (1998) [13]. Both its consistency and its asymptotic normality are precised by Mikosch and Straumann (2006) [19] who set a nice theoretical framework for the univariate case. For multivariate time series we defer the reader to Bardet and Wintenberger (2007) [1]. For GARCH models, mention among others the works of Lee and Hansen (1994) [17], Lumsdaine (1996) [18], Berkes, Horváth and Kokoszka (2003) [4] and Francq and Zakoïan (2004) [10]. As we will see, asymptotics properties of the smoothed QMLE can be obtained using the same arguments as for the classical QMLE.

Let us introduce some assumptions :

(A1) : $\gamma(A(\theta_0)) < 0$.

(A2) : The roots of the polynomial P defined by $P(z) = 1 - \sum_{j=1}^{q} b_{0,j} z^j$ are outside the unit disk.

(A3) : $\theta_0 \in \Theta$, a compact set such as for all $\theta \in \Theta$, the first component a_0 of θ is strictly positive.

(A4) : The support of the law of ξ_t admits more than 2 points.

(A5) : θ_0 belongs to the interior Θ° of Θ.

(A6) : $\mathbb{E}X_0^4 < \infty$.

The top-Lyapounov exponent $\gamma(A(\theta_0))$ is defined for the LARCH part only, as in (7.8). Assumptions **(A1)** and **(A2)** ensure existence and uniqueness of the AR-LARCH process (7.9). The two following results are devoted respectively to *a.s.* consistency and to the central limit behaviour of the smoothed QMLE.

Theorem 7.1 *Under assumptions* $(\mathbf{A1}) - (\mathbf{A4})$ *the smoothed QMLE is consistent for each value of* $h > 0$ *:*

$$\lim_{n\to\infty} \hat{\theta}_{n,h} = \theta_0, \qquad a.s.$$

Theorem 7.2 *If* **(A1)**-**(A6)** *hold true, the smoothed QMLE is asymptotically normal for each value of* $h > 0$ *:*

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \xrightarrow{\mathcal{D}}_{n\to\infty} \mathcal{N}\left(0, N_h^{-1} M_h N_h^{-1}\right).$$

where

$$\begin{aligned}
N_h &= N_h^{(1)} + N_h^{(2)}, \quad M_h = M_h^{(1)} + M_h^{(2)} + M_h^{(3)}, \\
N_h^{(1)} &= 2\mathbb{E}\left(\frac{\nabla m_0(\theta_0)\nabla m_0(\theta_0)'}{V_0(\theta_0)+h}\right), \quad N_h^{(2)} = \mathbb{E}\left(\frac{\nabla V_0(\theta_0)\nabla V_0(\theta_0)'}{(V_0(\theta_0)+h)^2}\right), \\
M_h^{(1)} &= 4\mathbb{E}\left(\frac{V_0(\theta_0)\nabla m_0(\theta_0)\nabla m_0(\theta_0)'}{(V_0(\theta_0)+h)^2}\right), \\
M_h^{(2)} &= (\mu_4 - 1)\mathbb{E}\left(\frac{V_0(\theta_0)^2\nabla V_0(\theta_0)\nabla V_0(\theta_0)'}{(V_0(\theta_0)+h)^4}\right), \\
M_h^{(3)} &= 2\mu_3\mathbb{E}\left(\frac{V_0(\theta_0)\sigma_0(\theta_0)}{(V_0(\theta_0)+h)^3}\left(\nabla m_0(\theta_0)\nabla V_0(\theta_0)' + \nabla V_0(\theta_0)\nabla m_0(\theta_0)'\right)\right).
\end{aligned}$$

Remark. If $q = 0$ then Y is a pure LARCH model (7.10) and we obtain the consistency and the asymptotic normality of the smoothed QMLE as above. Its asymptotic variance writes as $\left(N_h^{(2)}\right)^{-1} M_h^{(2)} \left(N\right.$

7.5 Choice of the smoothing parameter h

We aim here at precising the asymptotic variance of the smoothed QMLE when $h \to 0$. We denote by $\|\cdot\|$ the Euclidean norm for a vector or a matrix. For simplicity we write m (resp. V) instead of $m_0(\theta_0)$ (resp. $V_0(\theta_0)$) and $\nabla m, \nabla V$ for the gradient vectors.
Using the notations in Theorem 7.2 we denote $v_h = N_h^{-1} M_h N_h^{-1}$ the asymptotic variance of the smoothed QMLE (see Theorem 7.2).
Unexpected results appear by plotting the asymptotic variance of the smoothed QMLE for small values of h. Suppose that we want to estimate the parameter a of the model : $X_t = \xi_t(1 + aX_{t-1})$ where $\xi_0 \sim \mathcal{N}(0,1)$. Then the asymptotic variance of the smoothed QMLE denoted by $v_h(a)$ seems to verify $\lim_{h\to 0} v_h(a) = 0$ for a large subset of parameters (see figure 7.2).

To study the behaviour of the asymptotic variance we set $A \preceq B$, the relation of order between symmetric positive definite matrices such that $x'Ax \le x'Bx$ if for each $x \in \mathbb{R}^d$, here A and $B \in \mathcal{M}_d(\mathbb{R})$.
In the following Lemma we discuss the qualitative behaviour of $h \mapsto v_h$. Even if we were not able to check monotonicity of this function (with the order $\preceq$), we shall precise in some cases the behaviour

of the asymptotic variance at the origin : here $v = \lim_{h\to 0^+} v_h = \inf_{h>0} v_h$ is either degenerated or has the same form that the asymptotic variance of the classical QMLE.

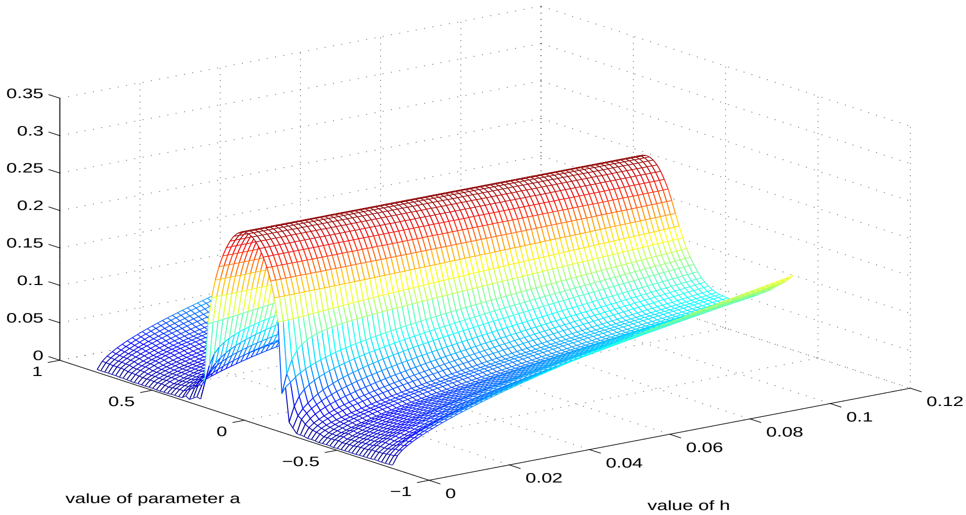


FIG. 7.2: $(h, a) \mapsto v_h(a)$.

The behaviour of the asymptotic variance near $h = 0$ is related to the condition :

$$\mathcal{C}: \quad \mathbb{E}\left(\frac{\|\nabla m\|^2}{V}\mathbb{1}_{(\nabla m, V)\neq(0,0)}\right) + \mathbb{E}\left(\frac{\|\nabla V\|^2}{V^2}\mathbb{1}_{V\neq 0}\right) < \infty.$$

Of course if $q = 0$ then the condition $\mathcal{C}$ reduces to

$$\mathbb{E}\left(\frac{\|\nabla V\|^2}{V^2}\mathbb{1}_{V\neq 0}\right) < \infty.$$

One can remark that when $\xi_0 \sim \mathcal{N}(0,1)$, by lemma 7.2 we have $V \neq 0$ a.s and condition $\mathcal{C}$ ensures the existence of the conditional Fisher information which is $\frac{1}{4}\text{Var}\ (\nabla q_{0,0}(\theta_0))$.

Lemma 7.3 *Let the assumptions of Theorem 7.2 hold.*

1. *If condition $\mathcal{C}$ does not hold, then $\lim_{h\to 0}\lambda_h = 0$, where λ_h is the smallest eigenvalue of v_h.*
2. *We suppose that either $q = 0$, or $q \neq 0$ but $(\mu_3, \mu_4) = (0, 3)$ (i.e ξ_0 has the same four first moments that a standard Gaussian). Then : $v = \lim_{h\to 0^+} v_h$ exists, and $v \preceq v_h$, for $h > 0$. Moreover v is non degenerate if and only if condition $\mathcal{C}$ holds. In this case, $v = (\mu_4 - 1)N^{-1}$ where :*
$$N = 2\mathbb{E}\left(\frac{\nabla m \nabla m'}{V}\mathbb{1}_{(\nabla m, V)\neq(0,0)}\right) + \mathbb{E}\left(\frac{\nabla V \nabla V'}{V^2}\mathbb{1}_{V\neq 0}\right)$$

Remarks. **a.** Condition $\mathcal{C}$ holds if there exists $m > 0$ with :

$$V_0(\theta_0) \geq m > 0 \qquad a.s. \tag{7.16}$$

This is the case for example if ξ_0 has a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$ and $a = \sum_{j=1}^{p}|a_{0,j}| < \frac{1}{2\sqrt{3}}$. Indeed we have $V_0(\theta_0) = a_{0,0}^2\left(1 + \widetilde{\sigma}_0\right)^2$ where

$$\widetilde{\sigma}_0(\theta_0) = \sum_{k\geq 1}\sum_{j_1,\ldots,j_k\in\{1,\ldots,p\}} a_{0,j_1}\cdots a_{0,j_k}\xi_{-j_1}\cdots\xi_{-(j_1+\ldots+j_k)}.$$

Note that $|\widetilde{\sigma}_0(\theta_0)| \leq \frac{a\sqrt{3}}{1-a\sqrt{3}} < 1$ and $m = a_{0,0}^2\left(1 - \frac{a\sqrt{3}}{1-a\sqrt{3}}\right)^2$ is a convenient value for (7.16) to hold. Condition (7.16) is a classical assumption to get the asymptotic properties of the classical QMLE, but for model (7.9), this kind of restriction seems unrealistic.

b. From point 1 in Lemma 7.3, if the condition $\mathcal{C}$ does not hold, then no asymptotically efficient estimator with $\sqrt{n}$-rate can be exhibited. This is the case if :

$$\mathbb{E}V_0(\theta_0)^{-1} \quad \mathbb{1}_{V_0(\theta_0)\neq 0} = \infty. \tag{7.17}$$

Condition (7.17) is related to the behaviour of $V_0(\theta_0)$ around 0. The following (artificial) example shows that this condition may hold.
Suppose that $b_{0,1} = \ldots = b_{0,q} = 0$ and $p = 1$, $a_{0,0} = 1$, $a_{0,1} = 0.5$, $\mathbb{P}(\xi_0 = 1) = \mathbb{P}(\xi_0 = -1) = \alpha$ and $\mathbb{P}(\xi_0 = 0) = 1 - 2\alpha$ for $\alpha \in [1/4, 1/2)$.
Then $\mathbb{E}\xi_0 = 0$ and from (7.7), $\mathbb{E}X_0^4 < \infty$ and we may assume $\mathbb{E}\xi_0^2 = 1$, for this we write $X_t = \xi_t/\sqrt{\mathbb{E}\xi_0^2}\left(\sqrt{\mathbb{E}\xi_0^2} + 0.5\sqrt{\mathbb{E}(\xi_0^2)}\, X_{t-1}\right)$. From (7.5) the chaotic expansion of the solution writes :

$$X_t = \xi_t + \sum_{j\geq 1} 2^{-j}\xi_t \cdots \xi_{t-j}$$

Let $n \in \mathbb{N}^*$ and suppose that $\xi_t = -1, \xi_{t-1} = \cdots = \xi_{t-n} = 1$ and $\xi_{t-(n+1)} = 0$, then :

$$X_t = -2(1 - 2^{-(n+1)})$$

and thus $V_{t+1}(\theta_0) = 2^{-(2n+2)}$. We now deduce :

$$\begin{aligned}
\mathbb{E}\frac{\mathbb{1}_{V_{t+1}(\theta_0)\neq 0}}{V_{t+1}(\theta_0)} &\geq \sum_{n\geq 1} 2^{2n+2}\mathbb{P}\left(V_{t+1}(\theta_0) = 2^{-(2n+2)}\right) \\
&\geq \sum_{n\geq 1} 2^{2n+2}\mathbb{P}\left(\xi_t = -1, \xi_{t-1} = \cdots = \xi_{t-n} = 1, \xi_{t-(n+1)} = 0\right) \\
&= (1-2\alpha)\sum_{n\geq 1}\alpha^{n+1}2^{2n+2} \\
&= \infty
\end{aligned}$$

This example shows that the condition (7.17) may happen to hold. Now figure 7.2 seems to prove that the model LARCH(1) also exhibits this condition for $\xi_0 \sim \mathcal{N}(0,1)$ but no formal proof is given here.

c. It is clear that both the QMLE and the smoothed QMLE will apply for classical ARCH models for which the conditional variance is bounded away from zero. In general the asymptotic variance will write as $v_h, h \geq 0$ where m, V denote the two first conditional moments of the process. Then the proof of the point 2 of lemma 7.3 shows that the QMLE is more efficient then the smoothed QMLE.

7.6 Numerical illustration

We illustrate the behaviour of the smoothed QMLE with an example. Our goal here is to see if $h \to 0$ gives best estimators as suggested by the Lemma 7.3. We set $p = q = 1$ and we consider Gaussian errors. We recall that asymptotic normality of the smoothed QMLE requires $\mathbb{E}Y_0^4 < \infty$ (Theorem 7.2) ; moreover $\mathbb{E}X_0^4 < \infty \Rightarrow \mathbb{E}Y_0^4 < \infty$. The following Lemma gives a necessary and sufficient condition for the existence of the fourth moment of the solution of

$$X_t = \xi_t\left(a_{0,0} + a_{0,1}X_{t-1}\right) \tag{7.18}$$

if $\mathbb{E}\xi_0^3 = 0$.

Lemma 7.4 *Suppose that* $\mathbb{E}\xi_0^3 = 0$ *then there exists a stationary solution of equation (7.18) with* $\mathbb{E}X_0^4 < \infty$ *if and only if* $a_1^4 \mathbb{E}\xi_0^4 < 1$. *In this case this solution is the unique stationary solution of equation (7.18).*

Remarks.

- If $\xi_0 \sim \mathcal{N}(0,1)$ is a standard normal random variable the condition $a_{0,1}^4 \mathbb{E}\xi_0^4 < 1$ writes $|a_{0,1}| < 3^{-1/4} \approx 0.7598\cdots$.
- If ξ_0 follows the uniform distribution law on the interval $[-\sqrt{3}, \sqrt{3}]$ this condition writes $|a_{0,1}| < (5/9)^{1/4} \approx 0.8633\cdots$. Thus if $|a_{0,1}| > 1/\sqrt{3} \approx 0.5774\cdots$, the process X is not bounded.

For the simulation study we have computed 500 smoothed QML estimators for sample sizes $n = 100$ and $n = 1000$ and for the smoothing parameters $h = 0.5, 0.1$ and 0.001. The value of the parameter is $\theta_0 = (0.5, 1.6, -0.7)$.

An expected problem is the irregularity of the function $Q_{n,h}$ when h is small. This holds even for very large values of n. As an example we plot $a_1 \mapsto Q_{n,h}(a_1)$ in figure 7.3 for the model :

$$X_t = \xi_t \left(1 + 0.5X_{t-1}\right), \qquad \xi_0 \sim \mathcal{N}(0,1).$$

Then, to avoid optimization problems, we first compute the estimators for $h = 0.5$; after this, using those values to initialize the procedure, we start with an optimization procedure for $h = 0.1, 0.001$. We see from figure 7.4 that the mean square errors decrease as soon as h decreases. However if h is small, fitting to a Gaussian distribution is not very good for $n = 100$ (figure 7.5 and figure 7.7) but a large sample size $n = 1000$ corrects this problem (figure 7.6 and figure 7.8). Hence the choice of the value of $h = h_n$ (depending on the sample size n) seems crucial. This problem is beyond the scope of this paper because we did not exhibit a balance of terms explaining this phenomenon as this is usual *e.g.* for non-parametric estimation.

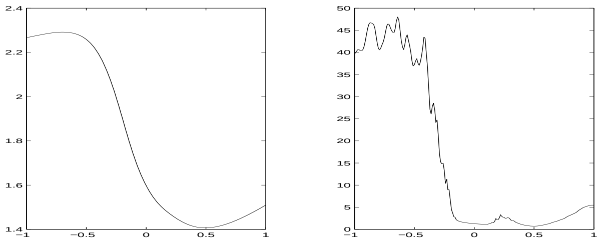


FIG. 7.3: $Q_{n,h}$, $n = 20000$, $h = 0.5$ or $h = 0.001$.

Estimators	Sample size	S. QML		
		$h = 0.5$	0.1	0.001
$\hat{b}_1$	$n = 100$	1.9	1.1	0.9
$\hat{a}_0$	$n = 100$	16.3	14.7	13.9
$\hat{a}_1$	$n = 100$	8	5.8	5.1
$\hat{b}_1$	$n = 1000$	0.1	0.1	0
$\hat{a}_0$	$n = 1000$	1.7	1.5	1.4
$\hat{a}_1$	$n = 1000$	0.6	0.4	0.3

FIG. 7.4: Mean square errors for the three estimators ($\times 10^{-3}$).

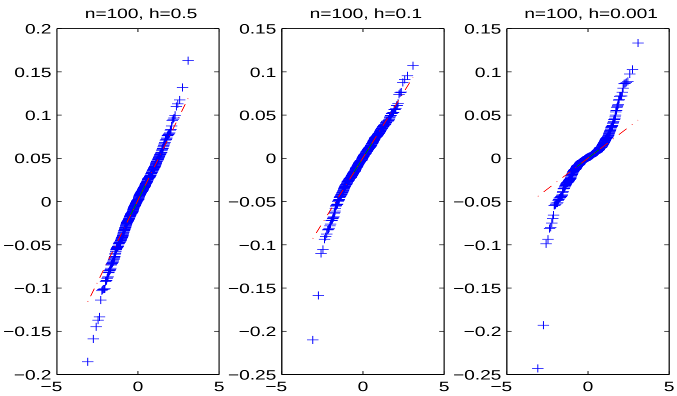


FIG. 7.5: Normal Q-Q plots for the errors $\hat{b}_1 - b_{0,1}$.

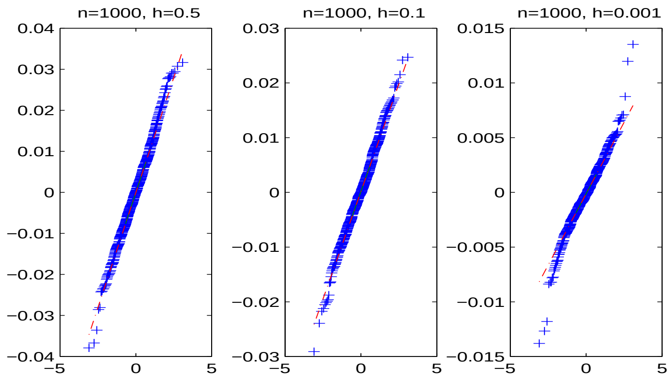


FIG. 7.6: Normal Q-Q plots for the errors $\hat{b}_1 - b_{0,1}$.

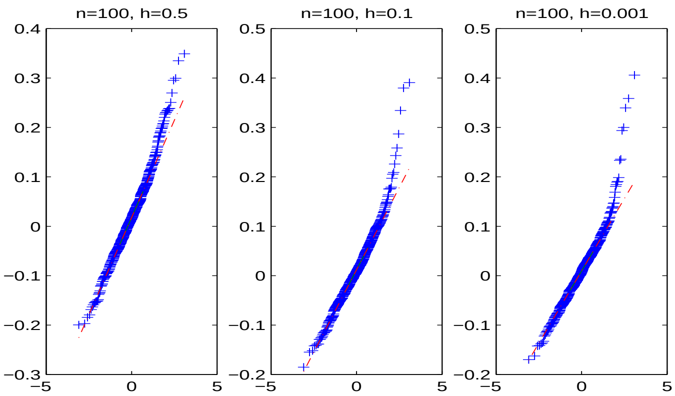


FIG. 7.7: Normal Q-Q plots for the errors $\hat{a}_1 - a_{0,1}$.

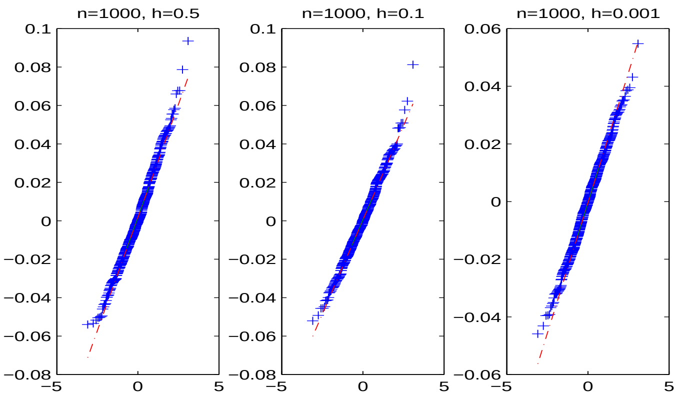


FIG. 7.8: Normal Q-Q plots for the errors $\hat{a}_1 - a_{0,1}$.

7.7 Proofs

7.7.1 Proof of Lemma 7.1

Let $A = \{(z_1, \dots, z_p) \in \mathbb{R}^p / a_{0,0} + \sum_{j=1}^p a_{0,j} z_j \neq 0\}$ and for $t \in \mathbb{Z}$, $Z_t = (X_{t-1}, \dots, X_{t-p})$. Set $\sum_{j=j_0+1}^p = 0$ if $j_0 = p$. Then since $a_{0,0} \neq 0$, $X_t \equiv 0$ is not a solution of the equation (7.10) thus $\mathbb{P}(Z_t \in A) > 0$. Moreover for $z \in A$ the support of the conditional law $\mathcal{L}(X_{-j_0}/Z_{-j_0} = z)$ is the whole set $\mathbb{R}$ which is in contradiction with the existence of a number $m > 0$ such that

$$V_0(\theta_0) = \left(a_{0,0} + a_{j_0} X_{-j_0} + \sum_{j=j_0+1}^{p} a_j X_{-j} \right)^2 \geq m, \quad \text{a.s.} \quad \square$$

7.7.2 Proof of Lemma 7.2

For simplicity, we denote σ_t instead of $\sigma_t(\theta_0)$. The result is obvious if $a_{0,j} = 0$, $j = 1, \dots, p$ since in this case $\sigma_0 = a_{0,0} \neq 0$.
Now let $j_0 \geq 1$ be the first index such that $a_{0,j_0} \neq 0$. Let $\alpha = \mathbb{P}(\sigma_0 = 0)$. We prove by induction that :

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(A_n) = \alpha \tag{7.19}$$

where for $n \in \mathbb{N}$, $A_n = \bigcap_{l=0}^{n} \{\sigma_{-lj_0} = 0\}$. This will conclude the proof. Indeed setting $n \to \infty$, we deduce that $\mathbb{P}\left(\bigcap_{l=0}^{\infty} \{\sigma_{-lj_0} = 0\} \right) = \alpha$ and from the ergodicity of the process $(\sigma_{-lj_0})_{l \in \mathbb{N}}$ we derive that $\alpha \in \{0, 1\}$. However $\alpha = 1$ implies by definition $\sigma_0 = X_0 = 0$ *a.s* which is impossible if $a_{0,0} \neq 0$: hence $\alpha = 0$.
We now prove (7.19). The definition of α implies the result for $n = 0$. Suppose that $\mathbb{P}(A_n) = \alpha$ and let us prove that $\mathbb{P}(A_{n+1}) = \alpha$. Then it is enough to prove $\mathbb{P}\left(A_n \cap \{\sigma_{-(n+1)j_0} \neq 0\}\right) = 0$ or :

$$\mathbb{P}\left(\sigma_{-nj_0} = 0, \sigma_{-(n+1)j_0} \neq 0\right) = \mathbb{P}(\sigma_0 = 0, \sigma_{-j_0} \neq 0) = 0. \tag{7.20}$$

Now $\sigma_0 = 0 \Leftrightarrow X_{-j_0} = \xi_{-j_0} \sigma_{-j_0} = M$, with $M = -\left(\sum_{j=j_0+1}^p a_{0,j} X_{-j}\right)/a_{0,j_0}$ (by convention $\sum_{j=p+1}^p = 0$). If μ is the law of the vector (σ_{-j_0}, M), we get using independence :

$$\mathbb{P}(\sigma_0 = 0, \sigma_{-j_0} \neq 0) = \int \mathbb{P}(a\xi_{-j_0} = b, a \neq 0) \mu(da, db).$$

Since $\mathbb{P}(\xi_0 = x) = 0$, $\forall x \in \mathbb{R}$, we have etablished (7.20), and (7.19) follows by induction. $\square$

7.7.3 Proof of Theorem 7.1

We first prove the following Lemma, useful to show that the parameter θ_0 is identifiable in the model (7.9).
Here Y is a model satisfying (7.9) and note that $\mathcal{F}_t = \sigma\left(X_{t-j}/j \in \mathbb{N}\right) = \sigma\left(Y_{t-j}/j \in \mathbb{N}\right)$ for $t \in \mathbb{Z}$.

Lemma 7.5 *We suppose that* **(A3)** *holds. Let U_1 and U_2 be two random variables measurable w.r.t $\mathcal{F}_{-1}$ and $(\alpha_j)_{0\le j\le p}$ and $(\beta_j)_{0\le j\le p}$ be real numbers such that $\beta_0 \neq 0$. Then*

1. $(X_0 - U_1) \times U_2 = 0$ *a.s* $\Rightarrow U_2 \times \sigma_0(\theta_0) = 0$ *a.s and* $U_1 \times U_2 = 0$ *a.s.*
2. $\mathbb{P}\left((X_0 - U_1) \times (X_0 - U_2) = 0\right) < 1.$
3. $\mathbb{P}\left(\sigma_0(\theta_0)\left(\beta + \sum_{j=1}^{p} \beta_j X_{-j}\right) = 0\right) < 1.$
4. $\left(\alpha_0 + \sum_{j=1}^{p} \alpha_j X_{-j}\right) \times \left(\beta_0 + \sum_{j=1}^{p} \beta_j X_{-j}\right) = 0,$ *a.s implies* $\alpha_j = 0$, *for all* $j = 0, \ldots, p$.

Proof.

1. Here $(X_0 - U_1)\, U_2 = 0$ *a.s* $\Rightarrow U_2\sigma_0(\theta_0)\xi_0 = U_1 U_2$. Since ξ_0 is not a constant and it is independent of $(U_2\sigma_0(\theta_0), U_1U_2)$ we have $U_2\sigma_0(\theta_0) = 0$ *a.s.*, thus obviously $U_1U_2 = 0$ *a.s.*
2. If $(X_0 - U_1)\,(X_0 - U_2) = 0$ *a.s* then

$$\sigma_0^2(\theta_0)\xi_0^2 + \sigma_0(\theta_0)\left(-U_1 - U_2\right)\xi_0 = -U_1U_2 \quad a.s$$

   Since $a_{0,0} \neq 0$, we have $X \neq 0$ a.s. Then there exists a realization of $X_{-1}, X_{-2}, \ldots$ such that $\sigma_0(\theta_0) \neq 0$. For such a realization the support of the conditional law $\xi_0/X_{-1}, X_{-2}, \ldots$ (by independence this is also the law of ξ_0) contains only two values. This yields a contradiction with **(A3)** and the result follows.
3. We suppose

$$\sigma_0(\theta_0)\Big(\beta_0 + \sum_{j=1}^{p} \beta_j X_{-j}\Big) = 0, \quad a.s \tag{7.21}$$

   We set $\beta_j = 0$ for $j \geq p+1$. Suppose that $l = \inf\{i \geq 1/a_{0,i} \neq 0\}$ exists. We will show by induction that for all $i \in \mathbb{N}$:

$$\sigma_{-il}(\theta_0)\Big(\beta_0 + \sum_{j\geq il+1} \beta_j X_{-j}\Big) = 0, \qquad a.s.$$

   From (7.21), the result holds for $i = 0$. Suppose that for $i \in \mathbb{N}$:

$$\sigma_{-il}(\theta_0)\Big(\beta_0 + \sum_{j\geq il+1} \beta_j X_{-j}\Big) = 0 \qquad a.s.$$

Then successive applications of point 1) lead to

$$\sigma_{-il}(\theta_0)\Big(\beta_0 + \sum_{j\geq(i+1)l} \beta_j X_{-j}\Big) = 0, \quad a.s.$$

Now as $\sigma_{-il}(\theta_0) = a_{0,0} + \sum_{j=l}^{p} a_{0,j}X_{-il-j}$ and $a_{0,l} \neq 0$ we deduce from point 2) that $\beta_{(i+1)l} = 0$. Moreover a new application of point 1) leads to $\sigma_{-(i+1)l}(\theta_0)\left(\beta + \sum_{j\geq(i+1)l+1} \beta_j X_{-j}\right) = 0$ *a.s.* Hence the result follows by induction.
Now if i is large enough :

$$\sigma_{-il}(\theta_0)\beta_0 = 0 \quad a.s.$$

Since $\beta_0 \neq 0$ this implies $\sigma_{-il}(\theta_0) = 0$ a.s. We have obtained a contradiction since X cannot equals 0 when $a_{0,0} \neq 0$.
If now l does not exist, we have $X = a_{0,0}\xi$ and equation (7.21) becomes $a_{0,0}\left(\beta_0 + \sum_{j\geq1} \beta_j a_{0,0}\xi_{-j}\right) = 0$ *(a.s)*. Taking expectations this equality leads to $\beta_0 = 0$ and we thus exhibit a contradiction. We have shown that relation (7.21) is not possible and the result follows.

4. Setting $\alpha_j = \beta_j = 0$ if $j \geq p+1$ we suppose $\left(\alpha_0 + \sum_{j\geq1} \alpha_j X_{-j}\right) \times \left(\beta_0 + \sum_{j\geq1} \beta_j X_{-j}\right) = 0$ *(a.s.)* An application of point 2) implies $\alpha_1 \times \beta_1 = 0$. Moreover an application of point 1) gives

$$\Big(\alpha_0 + \sum_{j\geq2} \alpha_j X_{-j}\Big) \times \Big(\beta_0 + \sum_{j\geq2} \beta_j X_{-j}\Big) = 0$$

Then by an induction argument, it is obvious that for $i \geq 1$, we will obtain $\alpha_i\beta_i = 0$ and

$$\Big(\alpha_0 + \sum_{j\geq i} \alpha_j X_{-j}\Big) \times \Big(\beta_0 + \sum_{j\geq i} \beta_j X_{-j}\Big) = 0 \quad a.s. \tag{7.22}$$

With $i \to \infty$ we thus derive $\alpha_0\beta_0 = 0$, hence $\alpha_0 = 0$. Suppose that there exists some $i \in \mathbb{N}^*$ such that $\alpha_i \neq 0$. Then $\beta_i = 0$ and applying point 1) to equality (7.22) we get

$$\sigma_{-i}(\theta_0)\Big(\beta_0 + \sum_{j\geq i+1} \beta_j X_{-j}\Big) \qquad a.s.$$

We obtain a contradiction using the stationarity and the point 3) considered with β_{j+i} instead of β_j, $j \geq 1$. Hence $\alpha_i = \alpha_0 = 0$, for all $i \geq 1$ and the result follows. $\square$

From the previous Lemma we deduce the identification condition :

Lemma 7.6 *If* **(A3)** *holds then*

$$(m_0(\theta), V_0(\theta)) = (m_0(\theta_0), V_0(\theta_0)) \quad a.s \Rightarrow \theta = \theta_0$$

Proof. The equality $m_0(\theta) = m_0(\theta_0)$ writes as :

$$\sum_{j=1}^{q}(b_j - b_{0,j})Y_{t-j} = 0 \quad \text{a.s.}$$

If there exists $j \in \{1, \ldots, q\}$ such that $b_j \neq b_{0,j}$, then there exists $j \in \{1, \ldots, q\}$ such that $X_{t-j} = \xi_{t-j}\sigma_{t-j}(\theta_0) \in \mathcal{F}_{t-j-1}$. Using the same argument as in the proof of point 1. in Lemma 7.5, we obtain $\sigma_{t-j}(\theta_0) = 0$ a.s. Then $X = 0$ a.s and this is a contradiction with $a_{0,0} \neq 0$. We deduce that $b_j = b_{0,j}$, $\forall j \in \{1, \ldots, q\}$.
Assume that equality $V_0(\theta) = V_0(\theta_0)$ a.s holds, then as

$$m_0(\theta) = m_0(\theta_0) \quad \text{a.s} \Rightarrow m_{-j}(\theta) = m_{-j}(\theta_0), \quad j = 1, \ldots, p \quad \text{a.s},$$

we obtain using equation (7.9) :

$$\Big(a_0 + a_{0,0} + \sum_{j=1}^{p} (a_j + a_{0,j})\, X_{-j}\Big)\Big(a_0 - a_{0,0} + \sum_{j=1}^{p} (a_j - a_{0,j})\, X_{-j}\Big) = 0 \qquad \text{a.s}$$

As $a_0 + a_{0,0} > 0$ we obtain $a_0 = a_{0,0}$ by using point 4) in Lemma (7.5), and $a_j = a_{0,j}$ for all $j = 1, \ldots, p$. Thus $\theta = \theta_0$. $\square$
Now we prove theorem 7.1. The proof follows the proof of theorem 2.1 in Jeantheau [13] who proved the consistency of the QMLE for general multivariate ARCH models (see theorem 2.1 of that paper). As in [13] we use the following Theorem which is a staightforward generalisation of Theorem 1.12 in Pfanzagl (1969) for i.i.d data.

Theorem 7.3 *Let $(Y_t)_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic process, θ a parameter in Θ a compact of $\mathbb{R}^d$, and for $n \in \mathbb{N}^*$, Q_n be a contrast such that*

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} f\left(Y_t, \ldots, Y_{t-p}; \theta\right)$$

where f is a measurable function with real values and continuous in θ. Suppose that
1) $\mathbb{E}\inf_{\theta\in\Theta} f\left(Y_0, \ldots, Y_{-p}; \theta\right) > -\infty$.
2) $\theta \mapsto \mathbb{E}f\left(Y_0, \ldots, Y_{-p}; \theta\right)$ has a unique finite minimum at θ_0,
The minimum contrast estimator $\hat{\theta}_n$ associated to Q_n is thus strongly consistent : $\lim_{n\to\infty}\hat{\theta}_n = \theta_0$ a.s.

We apply Theorem 7.3 setting $f\left(Y_0, \ldots, Y_{-p}; \theta\right) = q_{0,h}(\theta)$. Obviously f is continuous in θ.

- Since $\inf_{\theta\in\Theta} f\left(Y_0, \ldots, Y_{-p}; \theta\right) \geq \ln h$, assumption 1) of Theorem 7.3 holds for the AR-LARCH process Y.

– We next prove that assumption 2) holds. We first prove that $Q_h(\theta_0) = \mathbb{E}q_{0,h}(\theta_0) \in \mathbb{R}$ (from the last point we know that $Q_h(\theta_0)$ is well defined and $\in \mathbb{R} \cup \{\infty\}$. From **(A1)**, Francq and Zakoïan [11] prove that $\mathbb{E}|X_0|^s < \infty$ for some $s \in (0,1]$ (see the proof of Theorem 4.2 in [11]). hence :

$$Q_h(\theta_0) = 1 + \frac{1}{s}\mathbb{E}\ln\left(V_0(\theta_0)+h\right)^s \leq 1 + \frac{h^s}{s} + \mathbb{E}V_0(\theta_0)^s < \infty.$$

Now we prove that for $\theta \in \Theta$, $Q_h(\theta) \geq Q_h(\theta_0)$ and the equality holds only when $\theta = \theta_0$.
For $\theta \in \Theta$, we have :

$$\mathbb{E}\left(\frac{(Y_0 - m_0(\theta))^2 + h}{V_0(\theta)+h}\right) = \mathbb{E}\left((A\xi_0 + B)^2 + C\right),$$

where $A = (V_0(\theta)+h)^{-1/2}\sigma_0(\theta_0)$, $B = (V_0(\theta)+h)^{-1/2}(m_0(\theta_0) - m_0(\theta))$ and $C = (V_0(\theta)+h)^{-1}h$. If μ is the law of the vector (A,B,C), then we obtain using independence properties :

$$\mathbb{E}\left((A\xi_0 + B)^2 + C\right) = \int \mathbb{E}\left((a\xi_0 + b)^2 + c\right)\mu(da,db,dc) = \mathbb{E}\left(A^2 + B^2 + C\right),$$

and we have proved that

$$Q_h(\theta) = \mathbb{E}\left(\frac{(m_0(\theta) - m_0(\theta_0))^2 + h}{V_0(\theta)+h} + \ln(V_0(\theta)+h)\right).$$

We obtain :

$$Q_h(\theta) - Q_h(\theta_0) = \mathbb{E}\frac{V_0(\theta_0)+h}{V_0(\theta)+h} - \ln\frac{V_0(\theta_0)+h}{V_0(\theta)+h} - 1$$

Since $\left(x - \ln x \geq 1,\ \forall x > 0\right)$ and $\left(x - \ln x = 1 \Leftrightarrow x = 1\right)$ we derive $Q_h(\theta_0) \leq Q_h(\theta)$ and :

$$Q_h(\theta) = Q_h(\theta_0) \Rightarrow m_0(\theta) = m_0(\theta_0), \quad V_0(\theta) = V_0(\theta_0) \quad a.s$$

From Lemma 7.6, we get $\theta = \theta_0$ which proves that assumption 2) of Theorem 7.3 holds.
Then the consistency of the smoothed QMLE follows from Theorem 7.3. □

7.7.4 Proof of Theorem 7.2

We use very classical arguments, the approach of Straumann [21] allows to derive an uniform law of the large numbers namely we will use :

Theorem 7.4 (Straumann (2006), Theorem 2.2.1, [21]) *Let $(v_t)_{t\in\mathbb{Z}}$ be a stationary ergodic sequence with values in $\mathcal{C}\left(K,\mathbb{R}^k\right)$, the space of real continuous functions on a compact $K \subset \mathbb{R}^d$. Assume $\mathbb{E}\sup_{\theta\in K}\|v_0(\theta)\| < \infty$ then :*

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n} v_t(\theta) - \mathbb{E}v_0(\theta)\right| = 0, \qquad a.s$$

Before we prove Theorem 7.2 we need the two following lemmas :

Lemma 7.7 *Suppose that* **A3** *holds and let* $c \in \mathbb{R}^{p+q+1}$ *such that* $c'\nabla m_0(\theta_0) = c'\nabla V_0(\theta_0) = 0$ *a.s. Then* $c = 0$.

Proof. We compute $\partial m_0/\partial b_j(\theta_0) = Y_{-j}$ for $j = 1, \dots, p$, $\partial V_0/\partial a_0(\theta_0) = 2\sigma_0(\theta_0)$ and $\partial V_0/\partial a_j(\theta_0) = 2X_{-j}\sigma_0(\theta_0)$ for $j = 1, \dots, p$. Suppose that there exists $c = (\mu_1, \dots, \mu_q, \lambda_0, \dots, \lambda_p) \in \mathbb{R}^{p+q+1}$ such that,

$$c'\nabla m_0(\theta_0) = c'\nabla V_0(\theta_0) = 0, \quad \text{a.s.}$$

Then we obtain :

$$\sum_{j=1}^{q} \mu_j Y_{-j} = 0, \quad \text{a.s.}$$

As for the proof of Lemma 7.6 we obtain $\mu_1 = \ldots = \mu_q = 0$.
Hence equality $c'\nabla V_0(\theta_0) = 0$ rewrites :

$$\sigma_0(\theta_0)\Big(\lambda_0 + \sum_{j=1}^{d} \lambda_j X_{-j}\Big) = 0 \qquad a.s.$$

As $a_{0,0} \neq 0$ an application of point 4) of Lemma 7.5 implies $\lambda_j = 0$, $j = 0, \dots, p$. We have shown that $c = 0$. $\square$
For the proof of theorem 7.2, the following moment conditions will be used :

Lemma 7.8 *If* $\mathbb{E}X_0^4 < \infty$ *then* $\mathbb{E}\left\|\nabla q_{0,h}(\theta_0)\right\|^2 < \infty$, $\mathbb{E}\sup_{\theta\in\Theta}\left\|\nabla^2 q_{0,h}(\theta)\right\| < \infty$.

Proof. We first notice that

$$\nabla q_{0,h}(\theta) = \frac{-2(Y_0 - m_0(\theta))\nabla m_0(\theta)}{V_0(\theta) + h} + \frac{\nabla V_0(\theta)\left(V_0(\theta) - (Y_0 - m_0(\theta))^2\right)}{\left(V_0(\theta) + h\right)^2} \tag{7.23}$$

$$\begin{aligned}
\nabla^2 q_{0,h}(\theta) &= \frac{2}{V_0(\theta) + h}\nabla m_0(\theta)\nabla m_0(\theta)' \\
&- \frac{2(Y_0 - m_0(\theta))}{V_0(\theta) + h}\nabla^2 m_0(\theta) \\
&+ \frac{2(Y_0 - m_0(\theta)}{(V_0(\theta) + h)^2}\left(\nabla m_0(\theta)\nabla V_0(\theta)' + \nabla V_0(\theta)\nabla m_0(\theta)'\right) \\
&+ \frac{V_0(\theta) - (Y_0 - m_0(\theta))^2}{(V_0(\theta) + h)^2}\nabla^2 V_0(\theta) \\
&+ \frac{h - V_0(\theta) + 2(Y_0 - m_0(\theta))^2}{(V_0(\theta) + h)^3}\nabla V_0(\theta)\nabla V_0(\theta)'
\end{aligned}$$

As Θ is bounded since it is compact, the expressions of σ_0 and m_0 for model 7.9 entails the existence of a real $K > 0$ such that :

$$\sup_{\theta\in\Theta}\left(|m_0(\theta)| + \|\nabla m_0(\theta)\| + |\sigma_0(\theta)| + \|\nabla\sigma_0(\theta)\| + \left\|\nabla^2\sigma_0(\theta)\right\|\right) \le U, \tag{7.24}$$

with $U = K\left(1 + \sum_{j=1}^{p+q}|Y_{-j}|\right)$. Moreover for model (7.9), $\nabla^2 m_0 = 0$.

– For $\nabla q_{0,h}$, we have :

$$\nabla q_{0,h}(\theta_0) = \frac{-2X_0\nabla m_0(\theta_0)}{V_0(\theta_0)+h} + \frac{\nabla V_0(\theta_0)(1-\xi_0^2)^2 V_0(\theta_0)}{(V_0(\theta_0)+h)^2}.$$

Then,

$$\|\nabla q_{0,h}(\theta_0)\|^2 \le \frac{4X_0^2U^2}{(V_0(\theta_0)+h)^2} + \frac{V_0(\theta_0)^3U^2(1-\xi_0^2)^2}{(V_0(\theta_0)+h)^4}$$

This leads to :

$$\mathbb{E}\,\|\nabla q_{0,h}(\theta_0)\|^2 \le \frac{3+\mu_4}{h}\mathbb{E}U^2.$$

As $\mathbb{E}X_0^4 < \infty \Rightarrow \mathbb{E}Y_0^4 < \infty \Rightarrow \mathbb{E}U^2 < \infty$, we obtain $\mathbb{E}\,\|\nabla q_{0,h}(\theta_0)\|^2 < \infty$.

– For the second assertion, using the definition of U and the inequality $\frac{1}{V_0(\theta)+h} \le \frac{1}{h}$, we see that the fourth first term in the expression of $\nabla^2 q_{0,h}(\theta)$ can be bounded uniformly with respect to θ by polynomials of degree four in the variables $|Y_0|, |Y_{-1}|, \ldots, \left|Y_{-(p+q)}\right|$. For the last term it is also the case since :

$$\begin{aligned}\frac{\left|h - V_0(\theta) + 2(Y_0 - m_0(\theta))^2\right|}{(V_0(\theta)+h)^3}\,\|\nabla V_0(\theta)\|^2 &\le \frac{h + U^2 + 2(|Y_0| + U)^2}{(V_0(\theta)+h)^3}4V_0(\theta)U^2 \\ &\le \frac{h + U^2 + 2(|Y_0| + U)^2}{h^2}4U^2.\end{aligned}$$

The result follows thus from $\mathbb{E}Y_0^4 < \infty$. $\square$

We now turn to the proof of Theorem 7.2. Since $\theta \in \Theta^\circ$, a Taylor expansion yields :

$$0 = \nabla Q_{n,h}(\hat\theta_{n,h}) = \nabla Q_{n,h}(\theta_0) + \widetilde{M}_n \cdot (\hat\theta_{n,h} - \theta_0),$$

with $\widetilde{M}_n(i,j) = \partial^2 Q_{n,h}(\gamma_i)/\partial\theta_i\partial\theta_j$, and $\left\|\hat\theta_{n,h} - \gamma_i\right\| \le \left\|\hat\theta_{n,h} - \theta_0\right\|$, for $i = 1, \ldots, p+q+1$. Hence,

$$-\sqrt{n}\nabla Q_{n,h}(\theta_0) = \sqrt{n}\,\widetilde{M}_n \cdot (\hat\theta_{n,h} - \theta_0) \tag{7.25}$$

For each $(\theta, t) \in \Theta \times \mathbb{Z}$, $\nabla^2 q_{t,h}(\theta)$ is a measurable function of $Y_t, \ldots, Y_{t-(p+q)}$, thus we infer that $\left(\nabla^2 q_{t,h}\right)_t$ is a stationary ergodic and $\mathcal{C}(\Theta, \mathbb{R}^{p+q+1} \times \mathbb{R}^{p+q+1})$-valued sequence. According to Lemma 7.8, $\sup_{\theta\in\Theta}\left\|\nabla^2 q_{0,h}(\theta)\right\|$ is an integrable random variable, hence from Theorem 7.4 :

$$\sup_{\theta\in\Theta}\left\|\nabla^2 Q_{n,h}(\theta) - \mathbb{E}\nabla^2 q_{0,h}(\theta)\right\| \to_{n\to\infty} 0.$$

From $\hat{\theta}_{n,h} \to_{n\to\infty} \theta_0$ *(a.s.)* we thus conclude

$$\widetilde{M}_n \to_{n\to\infty} N_h = \mathbb{E}\nabla^2 q_{0,h}(\theta_0), \qquad a.s. \tag{7.26}$$

Moreover N_h is non-singular ; indeed using expression (7.23), we have

$$N_h = 2\mathbb{E}\frac{\nabla m_0(\theta_0)\nabla m_0(\theta_0)'}{V_0(\theta_0)+h} + \mathbb{E}\frac{\nabla V_0(\theta_0)\nabla V_0(\theta_0)'}{(V_0(\theta_0)+h)^2},$$

and from Lemma 7.7 this matrix is positive-definite. Now,

$$\sqrt{n}\ \nabla Q_{n,h}(\theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla q_{n,h}(\theta_0), \quad \text{with } \mathbb{E}_{\mathcal{F}_{t-1}}\nabla q_0(\theta_0) = 0.$$

Since from Lemma 7.8, $\mathbb{E}\left\|\nabla q_{0,h}(\theta_0)\right\|^2 < \infty$, $\left(\nabla q_{t,h}(\theta_0)\right)_t$ is an ergodic stationary $\mathcal{F}_t$-martingale difference sequence with finite variance and from Theorem 23.1, page 206 in [5],

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla q_{t,h}(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, M_h), \quad \text{with } M_h = \operatorname{Var}\nabla q_{0,h}(\theta_0).$$

Thus we infer

$$\sqrt{n}(\hat{\theta}_{n,h} - \theta_0) \to_{n\to\infty} \mathcal{N}(0, N_h^{-1}M_hN_h^{-1})$$

Then the result follows from the expression of M_h which is easy to derive from the expression of $\nabla q_{0,h}$ (7.23) . $\square$

7.7.5 Proof of Lemma 7.3

1) We use the expression of v_h given in Theorem 7.2. Then, using the expression of $\nabla q_{0,h}(\theta_0)$ in (7.23) and the triangular inequality, we have
$M_h \preceq ((\mu_4 - 1)\vee 2)\, N_h$. This leads to the following inequality :

$$v_h \preceq ((\mu_4 - 1)\vee 2)\, N_h^{-1} \tag{7.27}$$

Then λ_h is smaller then $((\mu_4-1)\vee 2)\,\mu_h$, where μ_h is the smallest eigenvalue of N_h^{-1}. Then it is sufficient to show that $\lim_{h\to 0}\mu_h = 0$. As $\mu_h = 1/\rho(N_h)$ where $\rho(N_h)$ denotes the spectral radius of the matrix N_h, we need to show that $\lim_{h\to 0^+}\rho(N_h) = \infty$ or equivalently that $\lim_{h\to 0^+}\sum_{i=1}^{p+q+1} N_h(i,i) = \infty$. But

$$\sum_{i=1}^{p+q+1} N_h(i,i) = 2\mathbb{E}\|\nabla m\|^2(V+h)^{-1} + \mathbb{E}\|\nabla V\|^2(V+h)^{-2}$$

and with monotone convergence :

$$\lim_{h\to 0^+} 2\mathbb{E}\frac{\left\|\nabla m\right\|^2}{V+h} + \mathbb{E}\frac{\left\|\nabla V\right\|^2}{(V+h)^2} = 2\mathbb{E}\frac{\left\|\nabla m\right\|^2}{V}\mathbb{1}_{(\nabla m, V)\neq(0,0)} + \mathbb{E}\frac{\left\|\nabla V\right\|^2}{V^2}\mathbb{1}_{V\neq 0} = \infty.$$

Hence we conclude that $\lim_{h\to 0}\lambda_h = 0$.
2) We will only prove this point if $q \neq 0$ and $(\mu_3,\mu_4) = (0,3)$. The case $q = 0$ is omitted since its proof follows from straightforward modifications.
If $(\mu_3,\mu_4) = (0,3)$ we first remark that by (7.27), we have :

$$v_h \preceq 2N_h^{-1} \tag{7.28}$$

If $y, z \in \mathbb{R}^d$ and $h, k > 0$ we have :

$$\begin{aligned}\sqrt{2}\,y'N_k^{(1)}z &= 2\sqrt{2}\,\mathbb{E}\frac{(\nabla m/y)\times(\nabla m/z)}{V+k} \\ &\leq 2\sqrt{2}\,\mathbb{E}^{1/2}\left(\frac{(\nabla m/y)^2(V+h)}{(V+k)^2}\right)\times\mathbb{E}^{1/2}\left(\frac{(\nabla m/z)^2}{V+h}\right)\end{aligned}$$

With analogous arguments we also have :

$$\sqrt{2}\,y'N_k^{(2)}z \leq \sqrt{2}\,\mathbb{E}^{1/2}\left(\frac{(\nabla V/y)^2(V+h)^2}{(V+k)^4}\right)\times\mathbb{E}^{1/2}\left(\frac{(\nabla V/z)^2}{(V+h)^2}\right)$$

Now using from the inequality $(ac+bd)^2 \leq (a^2+b^2)(c^2+d^2)$:

$$2\left(y'N_kz\right)^2 \leq y'M_{k,h}y \times z'N_hz,$$

where $M_{k,h} = 4\mathbb{E}\frac{\nabla m\nabla m'(V+h)}{(V+k)^2} + 2\mathbb{E}\frac{\nabla V\nabla V'(V+h)^2}{(V+k)^4}$.
Now if $z = N_h^{-1}x$ and $y = N_k^{-1}x$, we get :

$$2x'N_hx \leq x'N_k^{-1}M_{k,h}N_k^{-1}x.$$

Since $\lim_{h\to 0} M_{k,h} = M_k$, we obtain using (7.28) :

$$\limsup_{h\to 0^+} x'v_hx \leq 2\limsup_{h\to 0^+} x'N_h^{-1}x \leq x'v_kx. \tag{7.29}$$

We deduce that

$$\limsup_{h\to 0^+} x'v_hx \leq \liminf_{k\to 0^+} x'v_k.$$

The last inequality is obviously an equality and we conclude that for $x \in \mathbb{R}^{p+q+1}$, $\lim_{h\to 0} x'v_hx$ exists and belongs to $\mathbb{R}^+$. By polarization $\lim_{h\to 0} x'v_hy$ exists for all $x, y \in \mathbb{R}^{p+q+1}$. Then one can define $v = \lim_{h\to 0^+} v_h$. From (7.29), we deduce $v \preceq v_k$ if $k > 0$.
Suppose now that $\mathcal{C}$ holds. Then from the dominated convergence theorem we prove that

$$\lim_{h\to 0} M_h = 2\lim_{h\to 0} N_h = 2N.$$

From Lemma 7.7 this limit is non degenerated. The expression of v in this case follows now from the continuity of the application $A \mapsto A^{-1}$.
Now if condition $\mathcal{C}$ does not hold, then the point 1 shows that v is degenerated.□

7.7.6 Proof of Lemma 7.4

If $a_{0,1}^4\mathbb{E}\xi_0^4 < 1$ then $a_{0,1}^2 < 1$ and from Theorem 2.1 in [14] there exists a unique stationary solution of equation (7.18). The fourth moment of this solution exists from (7.7).
If now there exists a stationary solution of equation (7.18) such that $\mathbb{E}X_0^4 < \infty$. As

$$\mathbb{E}X_0^4 = \mathbb{E}\xi_0^4\mathbb{E}\left(a_{0,0}^4 + a_{0,1}^4X_0^4 + 6a_{0,0}^2a_{0,1}^2X_0^2 + 4a_{0,0}a_{0,1}^3X_0^3 + 4a_{0,0}^3a_{0,1}X_0\right),$$

since $\mathbb{E}\xi_0^3 = 0$ implies $\mathbb{E}X_0^3 = 0$ we get :

$$\mathbb{E}X_0^4 = \mathbb{E}\xi_0^4\mathbb{E}\left(a_{0,0}^4 + a_{0,1}^4X_0^4 + 6a_{0,0}^2a_{0,1}^2X_0^2\right).$$

Hence $\left(1 - a_{0,0}^4\mathbb{E}\xi_0^4\right)\mathbb{E}X_0^4 = \mathbb{E}\xi_0^4\left(a_{0,0}^4 + 6a_{0,0}^2a_{0,1}^2\mathbb{E}X_0^2\right)$ and $a_{0,1}^4\mathbb{E}\xi_0^4 < 1$. $\square$

Bibliographie

[1] Bardet, J.-M., Wintenberger, O. (2007) *Asymptotic normality of the quasi maximum likelihood estimator for multidimensional causal processes.* Preprint `arXiv :0712.0679`.

[2] Beran, J., Schützner, M. (2008) *Estimation of dependence parameters for linear ARCH processes.* Working document.

[3] Black, F. (1976) *Studies in stock price volatility changes.* Proceedings of the Business and Economic Statistics Section, 177-181.

[4] Berkes, I., Horváth, L., Kokoszka, P. S. (2003) *GARCH processes : structure and estimation.* Bernoulli **9**, 201-227.

[5] Billingsley, P. (1968) *Convergence of Probability Measures.* John Wiley, New York.

[6] Doukhan, P., Oppenheim, G., Taqqu, M., *editors* (2003) *Theory and Applications of Long-range Dependence.* Birkhaüser, Boston.

[7] Doukhan, P., Teyssière, G., Winant, P. (2006) *Vector valued ARCH infinity processes.* Dependence in Probability and Statistics. Patrice Bertail, Paul Doukhan and Philippe Soulier Editors, Springer, New York.

[8] Engle, R. F. (1982) *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.* Econometrica **50**, 987-1008.

[9] Engle, R. F. (1990) *Stock volatility and the crash of Š87. Discussion.* The Review of Financial Studies **3**, 103-106.

[10] Francq, C., Zakoïan, J-M. (2004) *Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes.* Bernoulli, 10, 605-637.

[11] Francq, C., Makarova, S., Zakoïan, J-M. (2008) *A class of stochastic unit-root bilinear processes. Mixing properties and unit-root test.* Forthcoming in the Journal of Econometrics.

[12] Francq, C., Zakoïan, J-M. *Inconsistency of the QMLE and asymptotic normality of the weighted LSE for a class of conditionally heteroscedastic models.* Working document. http ://www.eea-esem.com/files/papers/EEA-ESEM/2008/1099/statbprocesses 7juillet08.pdf

[13] Jeantheau, T. (1998) *Strong consistency of estimators for multivariate ARCH models.* Econometric Theory **14**-1, 70-86.

[14] Giraitis, L., Robinson, P., Surgailis, D. (2000) *A model for long memory conditional heteroscedasticity.* Annals of Applied Probability **10**-3, 1002-1024.

[15] Giraitis, L., Leipus, R., Robinson, P., Surgailis, D. (2004) *LARCH, Leverage, and Long Memory.* Journal of Financial Econometrics **2**-2, 177-210.

[16] Giraitis, L., Surgailis, D. (2002) *ARCH-type bilinear models with double long memory.* Stochastic Processes and their Applications **100**, 275-300.

[17] Lee, S.-W., Hansen, B. E. (1994) *Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator.* Econometric theory **10**, 29-52.

[18] Lumsdaine, R. L. (1996) *Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models.* Econometrica **64**, 575-596.

[19] Mikosch, T., Straumann, D. (2006) *Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series : A stochastic recurrence equations approach.* Annals of Statistics **34**-5, 2449-2495.

[20] Sentana, E. (1995) *Quadratic ARCH models.* Review of Economic Studies **62**, 639-661.

[21] Straumann, D. (2004) *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics. Springer Verlag.

[22] Weiss, A. A. (1986) *Asymptotic theory for ARCH models : estimation and testing.* Econometric theory **2**, 107-131.